DOCUMENT RESUME

ED 037 805 24 CG 005 307

AUTHOR Rimoldi, H. J. A.; And Others
TITLE Problem Solving Processes Used by Elementary School Boys. Final Report.
INSTITUTION Loyola Univ., Chicago, Ill. Psychometric Lab.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-6-1089
PUB DATE Feb 70
GRANT OEG-3-6-061089-1020
NOTE 147p.

EDRS PRICE EDRS Price MF-$0.75 HC-$7.45
DESCRIPTORS *Ability, Cognitive Processes, *Elementary School Students, Language Proficiency, *Language Research, *Logical Thinking, Males, *Problem Solving, Structural Analysis

ABSTRACT

Boys (n=800) between the ages of seven and twelve were administered 35 problems for three successive years. Each problem was characterized according to its logical structure and language which corresponded to a pictorial presentation, its ordinary verbal language, and its abstract symbols. Each structure was presented in each of the three languages. Evaluation of subject performance was made using problem solving tactics. Results were treated by profile analysis. Subjects were classified by age, I.Q. levels and grade point averages. Contrasts examined on the basis of these classifications indicate that there is a differentiation between structures that increases with age. A general conclusion is that it is risky to rely exclusively on the final answer to a problem in appraising thinking ability, and that the differentiation between logical structures and languages is basic. (Author/CJ)

BR 6-1089
PA 24

C G

ED037805

Final Report
Project No. 6-1089
Grant No. OEG-3-6-061089-1020

2 C

PROBLEM SOLVING PROCESSES USED BY ELEMENTARY SCHOOL BOYS

H. J. A. Rimoldi, M.D., Ph.D.
T. W. Chlapecka
Mira Aghi, Ph.D.
Loyola Psychometric Laboratory
6525 N. Sheridan Road
Chicago, Illinois 60626

February, 1970

CG005307

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ED037805

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Final Report

Project No. 6-1089
Grant No. OEG-3-6-061089-1020

Problem Solving Processes Used By Elementary School Boys

H. J. A. Rimoldi, M.D., Ph.D.
T. W. Chlapecka
Mira Aghi, Ph.D.

Loyola Psychometric Laboratory
Loyola University
Chicago, Illinois

February, 1970

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

CG005307

Contents

Page
Acknowledgements. 3
Summary 4
Introduction. 6
Methods 26
Results 43
Conclusions and Recommendations 93
References 102
Appendix A 105
Appendix B 129

List of Tables

Table
1. Description of Testing Design with the Number of Subjects Tested Each Year. 27
2. Problem Identification. 29
3. General Methodological Design 39
4. General Hypotheses Concerning Structure and Language. . . 40
5. Summary, Means, Standard Deviations, and N. 63
6. Picture Problem Means, Standard Deviations, and Number for the Total Sample. 64
7. Verbal A Means, Standard Deviations, and Number for the Total Sample. 65
8. Verbal B Problem Means, Standard Deviations, and Number for the Total Sample. 66
9. Picture Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. . . 67
10. Picture Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. 68
11. Verbal A Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. 69
12. Verbal A Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. 70
13 Verbal B Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. 71
14. Verbal B Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups. 72
15. Picture Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups. . 73
16. Picture Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups. 74

Table Page

17. Verbal A Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium and Low G.P.A. Groups. 75
18. Verbal A Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups. 76
19. Verbal B Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups. 77
20. Verbal B Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups. 78
21. Multivariate Analysis Results: Comparisons of Different Age Groups. 79
22. Multivariate Analysis Results: Comparisons of High, Medium, and Low I.Q. Groups over the Total Sample. . . . 80
23. Multivariate Analysis Results: 7 Year Olds by I.Q. . . . 81
24. Multivariate Analysis Results: 8 Year Olds by I.Q. . . . 82
25. Multivariate Analysis Results: 9 Year Olds by I.Q. . . . 83
26. Multivariate Analysis Results: 10 Year Olds by I.Q. . . . 84
27. Multivariate Analysis Results: 11 Year Olds by I.Q. . . . 85
28. Multivariate Analysis Results: 12 Year Olds by I.Q. . . . 86
29. Multivariate Analysis Results: 7 Year Olds by G.P.A. . . 87
30. Multivariate Analysis Results: 8 Year Olds by G.P.A. . . 88
31. Multivariate Analysis Results: 9 Year Olds by G.P.A. . . 89
32. Multivariate Analysis Results: 10 Year Olds by G.P.A. . . 90
33. Multivariate Analysis Results: 11 Year Olds by G.P.A. . . 91
34. Multivariate Analysis Results: 12 Year Olds by G.P.A. . . 92

List of Figures

Figure

1. Illustration of Language Mappings. 21
2a. Structure 31. 22
2b. Structure 33. 22
2c. Structure 35. 22

Acknowledgements

The research presented in this report was supported by the Bureau of Research, United States Office of Education. It represents a natural outgrowth of studies that this office had supported in the past. Because of their continued help in all facets that pertain to a long range complex research problem and because of their permanent and wholehearted support, their welcome suggestions, and their patience, we are very deeply indebted.

In the 4-1/2 years that this research lasted, many persons were involved. The authors are thankful to all of them.

At the very early stages G. Burger, Ph.D. helped in planning the overall study. K. Vander Woude, Ph.D. was responsible for organizing the team of test administrators, preparing some of the problems and co-authoring the first progress report. Ana Maria Insua, Ph.D. prepared some of the problems used. Rev. I. Begin, S.J., Ph.D., Bernard Harney, Ph.D., Hugh Creedon, Margaret Sabol, Carol Jahn, and Bernard Clark contributed substantially to the testing of the subjects. We are indebted to Dr. Wozniak, Dean, School of Education, Loyola University, and Sister Stephenette, C.S.S.F. for making available the subjects used in this study.

For the extensive programming required in this study, we are thankful to Mr. D. Mazukelli and especially to Mr. Patrick McAuley, M.A.

Rev. T. Ferisin and Mr. Jeff Laack were responsible for preparing some of the tables and graphs accompanying this report. Miss Judy Heger was primarily responsible for all the secretarial work involved in this project. Her intelligent and conscientious approach went well beyond what is usually expected of a secretary's duties. Her help has been at all times invaluable.

The authors of this report want to express their sincerely felt gratitude for the generous contributions of all.

Summary

The purpose of this research was to study problem solving processes used by elementary school boys. Boys between the ages of 7 and 12 years were administered several problems for three successive years. The total number of subjects examined was 800 and 35 problems were prepared for this study.

Each problem administered was characterized considering its logical structure and its manner of presentation (language). The results refer to problems built using three structures and three languages. The languages correspond to a pictorial presentation (P), ordinary verbal language (VA), and abstract symbols (VB). Since these two components are operationally independent, each structure was presented in each one of the three languages. For retesting at successive years, parallel sets of problems were used.

The evaluation of a subject's performance was made considering his tactics in solving a problem. A tactic is defined by the sequence of questions that the subject asks in order to solve the problem. The experimenter provides the answers that correspond to the questions the subject asks. A general theoretical framework was developed and the relationship between tactics and processes was formulated. The subject's tactics are scored considering the structure of the problems. Specific consideration is given to order, redundancy, amount of information obtained, etc. The "pulling out" method of scoring used in this research considers these variables. Notice that the scoring method used is independent of sample evaluations. The results obtained were treated statistically by a multivariate approach, specifically profile analysis.

Subjects were classified in terms of age, I.Q. levels and grade point averages. On the basis of these classifications, several specific contrasts were examined. These refer to differentiation between structures, use of different languages, and effects of successive administrations on problem solving performance. These contrasts were studied for several possible combinations of problems, for instance various structures and different languages; same structure and same language.

The results indicate that there is a differentiation between structures that increases with age, but that at early ages this is better accomplished by problems in the P language. At around 9 to 10 years of age, language VA and VB become better discriminators. Several recommendations concerning psychological and educational problems have been made. A general conclusion is that in order to appraise thinking ability in problem solving performances, it is risky to rely exclusively on the final answer of a problem as it is usually done, and that the differen-.

tiation between logical structures and languages is basic. A subject may not be able to solve a problem, not because he does not understand its structure, but because he is not proficient in the use of the language in which it is presented. Thus in each individual case, thresholds for both logical structures and languages should be established.

It was further observed that in some cases discriminations between structures and between languages were reflected in I.Q. levels, or in G.P.A. levels. Specific examples are given in the text.

Introduction

General Considerations

The research to be reported in this volume concerns the experimental evaluation of problem solving tactics in children between seven and twelve years of age. The general design is longtitudinal since it studies the same subjects in three successive years. The results will be discussed with reference to age levels, independent evaluation of intelligence, and school grades.

The technique employed here to evaluate problem solving tactics departs from some of the approaches commonly used in this area of research. The major stress is in the experimentally controlled analysis of the process that subjects follow when solving a problem rather than in the exclusive evaluation of the final answer. Consequently the operational procedures used depart in several respects from those used in other investigations of problem solving and cognitive processes. These differences can be more readily understood by making explicit the assumptions that underlie our approach and the properties "built in" the instruments employed.

In many ways this research falls within the more inclusive chapter of cognitive studies. It may be properly qualified as epistemological in nature and development in design. Beyond studying independent groups of subjects at different age levels, we shall also experiment with the same subjects during a three year period. The findings that we shall report should complement those that we have obtained in many years of research conducted with independent samples of subjects at different ages and educational levels. The references to this work can be found at the end of this chapter. Indeed, some of the major theoretical considerations that substantiate this investigation resulted from this previous experimental work. Therefore this research is an independent testing of experimentally derived hypotheses.

Because of the type of experimentation that we have used, it is difficult to compare our results with those obtained by other investigators. Unless there is a reasonable communality in the assumptions made and in the operations performed, independent pieces of research may not always be successfully compared. As it is often the case, this results in endless argumentation after which no one gains necessarily in wisdom. In selecting the pertinent bibliography, we have limited the discussion to some of the recent work that is similar in assumptions, operations, and definitions to this study. If it is not always possible to conclude from similar results of independent investigations that a point has been proved or a hypothesis verified, then neither is it possible to conclude that different findings do represent divergent viewpoints. When in a restricted and a limited field of research sufficient theoretical and operational sophistication has been reached, some of these issues may be experimentally clarified by designing appropriate crucial experiments. Unfortunately, crucial experiments do not abound in science and they are not always easy to plan.

In order to obtain a better understanding of the issues involved in this research, we shall proceed by first characterizing some of the concepts used and second by discussing some of the theoretical and methodological issues involved in cognitive research. This will be followed by a statement of our assumptions, a description of the technique used, and the outline of the scoring procedures employed. Finally some theoretical and practical considerations will be presented with regard to the function played by logical structure and language in thinking processes.

Characterization of some terms and concepts used in this study

A cognitive process in a problem solving situation is defined as the sequence of psychological events and operations that occur when a subject solves a problem. This process begins with the presentation of a stimulus (problem or test item). The final solution (response) is an indication that the process has ended. A cognitive process is purposeful in the sense that it is directed towards attaining an end or goal or aim, more specifically the solution of the problem. This moving towards a goal confers a dynamic character to the process. In the realization of such directed activity the subject may aim at successive sub-goals, so that during the completion of the task (obtaining the solution) detours may occur.

The vicarious character of some cognitive processes is identified by the substitution of certain expected elements and relations by others that are functionally equivalent with regard to goal attainment.

Cognitive styles may be described in terms of the goal, sub-goals, detours, elements, and relations. No doubt environmental and genetic variables play a very basic role in determining the type of processes that subjects may preferentially follow. But these non-cognitive variables will not be explored in the present research.

For the experimental observation of processes it is necessary to provide the subject with a situation such that he will be free to pursue the attainment of the goal by bringing into play those elements that he considers necessary to solve a given problem. The laws of composition that hold in a given problem, that is, its structural properties, should be known prior to its administration. This allows a better understanding of the hunches of the hypotheses made and of how they are verified or rejected. It is obvious that in a concrete experimental situation it may be next to impossible to satisfy all these requirements. The technique used in this research satisfied some of them so that more important features of the subject's process could be observed.

Theoretical and methodological considerations involved in some cognitive research

Regardless of their philosophical leanings, psychologists have explicitly or implicitly stressed the dynamic character of the cognitive process. Some of these are Binet's experimental study of intelligence (1902), Spearman's formulation of his principles of cognition (1927), Thurstone's description of his primary mental ability factors (1938), Duncker's (1945), and Wertheimer's (1945), studies of problem solving, as well as Koehler's (1927) experiments, and those of Heibreder and Maier. Though the conclusions may differ, a common denominator that seems to run through all these researches refers to the dynamic character of cognitive performances and their purposefulness. A similar motive seems to run through more recent contributions as exemplified in Piaget's work, in Bartlett's experimentation, in Vigotski's approach, or in Bruner's strategies, to cite but a few. We shall look at some theoretical and methodological issues involved in the study of cognitive processes by analyzing them in terms of three major sub-groups: the experimental introspectionists, the testing approach and the observation of problem solving behavior as preferred by the Wertheimer group.

Besides the purely philosophical discussions, the experimental introspectionists have attempted to explore and characterize the ordered sequence of events implied in cognitive performance. In spite of the discredit that has unfortunately surrounded their efforts, it is fair to say that the deliberate effort made by these investigators to understand "what was going on in a subject's mind", provided some systematic knowledge. As a matter of fact some of these issues are still actively discussed. For example Binet's examination of the problem solving performance of his two daughters is still rich in heuristic value. But the difficulty of repeating their observations and the strong interaction between subject and experimenter has diminished their general acceptance.

With the advent and popularity of mental tests, these instruments became preferred tools for the investigation of thinking and problem solving and a new host of difficulties emerged. It is important to consider: (a) characterization and selection of test items or problems, (b) evaluation of responses, and (c) interpretation of test results.

A frequent approach to the characterization and selection of test items considers the proportion of subjects' responses that are right or wrong for a certain item. Seldom a discrimination is made according to the degree of correctness. On the basis of these results, items are either included or excluded in the final test form. This selection depends strongly on the sample used, so that for a different sample, different results may be expected. Subsequent test scores are dependent on these specific sample norms. At a later stage, the association between scores and independent criteria is determined to define the "thing that the test measures". It is not exceptional to discover after a period of time that a particular test is not associated with the criterion that it was meant to appraise originally.

For certain types of test items, it might be risky to use the procedures sketched above. For instance, let us take an item in which the subject has to give an answer to the problem: 2 is to 8, as 3 is to......There are many possible correct answers: for instance 9, (since 2 + 6 is equal to 8, and 3 + 6 equal to 9) or 12 (since 2 x 4 is equal to 8 and 3 x 4 is equal to 12), or 27 (since 2^3 is equal to 8, and 3^3 is equal to 27), or any even numbers (since 8 is even) or any odd number (since 2 and 8 are even but 3 is odd), or any number greater than 3 (since 8 is greater than 2), etc.

Obviously to score as correct the answer that is given with a certain predetermined frequency is not a criterion of correctness but only an indication of how much subjects of a given sample agree in the response they give. In a different sample, another answer may be more or less popular, but this doesn't make it either right or wrong. Niether does it make it right -- or wrong -- if in a large number of samples all the subjects would agree absolutely. Indeed, in the history of our civilization many a chapter deals with a reinterpretation or demonstration that some accepted correct answers or opinions were, after all, wrong. And in showing this some few people paid dearly.

Considering the score obtained by a subject, it is clear that the same performance may obtain different scores depending on the weights assigned to the items, since they in turn depend on the normative sample used. This affects the appraisal of a subject's ability and implies that the instrument is not invariant. Clearly the association that exists between scores and independent criteria will also change as a result of this lack of invariance.

This type of evaluation is of interest if we want to know how a subject stands in relation to a defined sample. But we may also be interested in knowing how to score the performance of a given subject regardless of the sample of subjects to which he belongs. This is more in line with measurement as applied in most of the physical sciences and in every day life situations where let us say, the meaning of X centimeters is the same regardless of whether the object measured is wood or steel or of any other material. In this case the ruler (instrument used for measurement) appraises specifically the property of the object in which we are interested, and in this sense it is a valid measurement. But it also implies that the properties of the ruler (test items, problems) are well known prior to its application.

The preceding paragraphs do not apply indiscriminately to all types of tests and are not meant to criticize the high level of ingenuity shown by many test theorists. In some cases the procedures that we have discussed are amply justified mainly in the case in which the test builder knows exactly what he wants to evaluate and how to do it. This refers essentially to instruments devised to appraise acquisition of certain specified types of knowledge, or achievement, etc. In these cases there is an attribute that we want to appraise, and since this attribute is (within limits) clearly defined, the construction of a specific test (valid) does not present an insurmountable difficulty. For instance if we are interested in appraising how much a subject knows about geography, no matter to which sample the subject belongs, saying that Latin America is north of the United States is wrong. But things are otherwise when the subject has to engender a response that cannot be evaluated exclusively in terms of authority, or previous knowledge, or a reordering of previously known things. For instance in the previous example: 2 is to 8 as 3 is to....., it is one thing to use this as a way of knowing whether the subject can multiply and divide successfully and another to see what kind of a process he uses to reach the answer. In the first case items directly involved in multiplication and division will probably be more satisfactory. But also 13 may be acceptable since 3 is odd and 2 and 8 are even. Or again 13 may be a good choice since (2 x 2) plus 4 = 8, and (3 x 3) + 4 = 13. Even if all subjects would answer 13, a perfectly objective answer, still we would not know how the subject reached it. Assuming that 13 is the only correct answer, we still do not know which process was followed by each individual to reach such a response. That is, the end product does not necessarily specify the inferences and the processes that preceded it. In other words, to know what has occurred between the presentation of the stimulus and the response may not always be safely inferred from the response.

Most of the studies of cognitive processes in which tests of reasoning, intelligence, general ability, problem solving, etc., are used, ignore this vicariousness of the process and rely almost exclusively on inferences made from the observed responses. Since the same response can be reached using different processes, then the inferences made are not as objective as more direct data on processes. And this state of affairs is little improved by refining the scoring methods or the statistical analysis or by multiplying the number of tests administered or by comparing the average performances of different samples, and so forth.

From the point of view of the psychology of individual differences, it is probable that by concentrating on the final answer we may be reducing our chances of discriminating between subjects. If it is at all true that the same answer can be reached in different ways, then the same final point can be reached by traveling along different paths. This convergence would reduce our possibility of differentiating between cognitive styles. In certain areas of knowledge it may be more important to know how a subject reaches a certain response than to know whether this response is or is not right.

However, it cannot be said that the study of final answers is irrelevant. The extreme opposite situation would correspond to the case in which everybody would think straight but nobody would give a correct answer. That is, solving a problem necessitates both the process and the solution, and a technique for the study of problem solving should ideally allow for the characterization of the process and the identification of the response. If the technique used sacrifices one of these two components, the results obtained are likely to be incomplete. In many practical situations, reaching a correct answer may be more important than following the best possible process.

With regard to the traditional type of problem solving situations, as analyzed by Duncker, Wertheimer, etc., the stress has been on the side of the process. These studies are in a way the counterpart of those conducted by investigators that use predominently the test approach and lack objectivity. The final response is not the main objective. On the contrary what subjects do and when they do it becomes the central issue. This is made experimentally observable by asking the subject to verbalize his thoughts, or by looking at the way in which the elements of the problem are ordered and classified by him and so forth. In some cases the elements that the subject has to consider in solving the problem are rather complex, for instance connecting strings, folding paper or transferring liquid from one receptacle to another, or restoring equilibrium, etc. Other times the problems are presented using drawings or ordinary language. In all cases, the problem solving situation terminates when the subject provides a final solution or elects not to proceed.

Looking at these experiments, it is not always possible to define exactly the elements and logical system of relations built in the problems. Do they, for example, deal with binary systems, the absorption law, or antisymmetry, etc.? In some cases this could be done and would provide a guiding principle in evaluating subjects' performances. On the other hand a more restrictive definition of the formal properties may be less fruitful than a free approach for general exploratory purposes. But if the structure of the problem is known and enough structures are explored, this shortcoming may be partially eliminated.

More disturbing is the fact that an interaction may exist between the manner of presentation and the structure of the problem. That is, we will not know whether not solving a problem is due to lack of understanding of the logical structure involved or to the subject's inability to operate with the concrete elements in which the problem has been presented. Concluding that the subject in question cannot handle the formal relations implied in the problem is an unwarranted assertion. But if the subject is given at least two problems with the same formal properties and different presentations (context) and if he is capable of solving one of them, then it may be concluded that the subject understands the structure involved in the problem.

It is indeed remarkable in fact that many every day problems and even routine performances may involve a complex network of relations and the subjects do this without major difficulty. But as soon as the same relational system is presented in abstract terms, few of them are able to solve it. The contrary situation also exists. The reasons why this occurs will not concern us in this monograph.

Another point to be considered with regard to problem solving experiments refers to the interaction that may exist between the subject and the experimenter. This could be controlled by rigorously defining what is expected of the experimenter in general and in unexpected situations. Sometimes conclusions are reached concerning cognitive performances in which these precautions are not strictly followed. It is then difficult to separate what is due to the subject and what is due to the tester. Though from a clinical point of view this may be acceptable, it is not so experimentally. During the actual testing, the subject's performance should not be either hampered or facilitated by extraneous remarks. Shifting the attention of the subject with helpful hints or proddings may be recommendable in diagnostic or clinical situations or in some type of learning experiments, but in our context identity of conditions should hold for all cases examined.

The results of these types of experiments are sometimes expressed in an anecdotal manner. Even though recording procedures give evidence of precision, it is still extremely difficult or even impossible to elaborate and generalize about the process involved.

In this context, Duncker's original study is exceptional considering the complexity of the situation involved. His efforts to differentiate what is observed from what is inferred at each step of the subject's or the experimenter's performance are well noted in the text. Unfortunately this is not always the case and the possibility of reproducing some modern studies in cognitive processes in a regorous experimental condition is, strictly speaking, very small.

Description of the technique used to evaluate cognitive processes in a problem solving situation

The essential feature of the technique consists of presenting a problem that the subject has to solve by asking a series of questions. This represents almost a reversal of the usual testing situation, in the sense that here the subject searches for questions that he thinks are appropriate to reach a solution. The experimenter "provides" a specific and fixed answer to each question asked. The subject is theoretically free to ask as many questions as he wants. Rather than being a passive receptor of stimuli, he is an active searcher of information. What he asks and when become features aiding an understanding of his process. Whenever the subject considers that he does not want to ask further questions, the testing situation is terminated.

The sequence of questions asked by the subject defines his tactic. Observable variables that characterize the tactic are the number, type, and order of questions asked. Further, since the solution offered is also available, it can be treated as a variable per se or in terms of its relation to the preceding sequence of questions asked.

The technique was originally used in 1954 to evaluate medical diagnostic skills, and has since given rise to other forms. The plasticity of the technique makes it readily applicable to a large variety of contexts for instance law cases, mathematical problems, everyday problems, problems in chemistry or biology, psychiatric diagnosis, Rorschach interpretations, etc. In over fifteen years of research the Loyola Psychometric Laboratory and others have developed hundreds of instruments using this approach.

In order to be consistent with our formulation of the problem, the subject should be able to ask any questions he wants whenever he wants. This poses some rather difficult problems for the experimenter in the sense that it is impossible to predetermine the whole universe of questions that may be asked and to generate a set of answers corresponding to such questions. Similarly if the experimenter is left free to answer the subject's questions there is always the possibility of a subject-experimenter interaction.

In order to avoid this, and to insure that a given question will always have the same answer we use the following procedure:

Each subject is presented a deck of 3 x 5 cards. On the first of these cards the problem is stated. Each of the remaining cards presents a question that the subject may ask. "Asking" a question means picking up a card. The pertinent answer is on the reverse side. Since each card is identified by a number, the experimenter or the subject can keep a record of the sequence of questions asked. Whether the problem is presented verbally or in writing, in some cases the possible questions correspond to drawings -- one drawing per card. For instance, assume that the problem consists of identifying the shape and color of a box, knowing that there are round and square boxes, yellow and blue. One of the cards (questions) may contain a drawing of round blue and round yellow boxes. The subject at a certain stage in the process may want to know if the box is round, in which case pointing to the indicated card he will receive the corresponding answer. If the experimenter says yes, this means that the box is round (but still the color is unknown); if the experimenter says no, then the subject should infer that the box has to be square. There are other possibilities for the mechanics of the asking and answering. For instance in Part III of the Test of the National Board of Medical Examiners (Hubbard, 1964), the subject obtains the answer by erasing a rectangle next to the question so that the response will appear.

The problems used in this research can be seen in Appendix A. In all cases the instructions indicate that the solution of the problem may be obtained by asking a certain series of questions. The subject is instructed to read the problem, examine all the questions and then to select only those he thinks are necessary for the solution.

An objection to the procedure just outlined is that subjects do not generate their own questions but choose them among those presented to them. In earlier studies (Rimoldi, Haley, Fogliatto, 1962a), we examined the limitations introduced by this modus operandi. In developing the test of medical diagnostic skills, every subject was asked to list separately those questions they would like to ask, but that were not presented with the problem. Judging from the results obtained, the set of questions given covered the area satisfactorily.

In another approach the subject was presented geometrical drawings in which he had to identify certain areas (Rimoldi, Devane, 1961a; Rimoldi, Fogliatto, Haley, Reyes, Erdmann, Zacharia, 1962b; Rimoldi, Fogliatto, Erdmann, Donnelly, 1964). This he could do by asking self-generated questions. This presentation will not be used in this study. Clearly the type of problem that can be thus presented is limited.

Assumptions made in the use of the technique

These assumptions have been discussed in previous studies (Rimoldi, 1967; Rimoldi, 1969) and are summarized below.

It is assumed that the tactic is an image of the process. To spell out the transformation that holds between the domain of processes and the codomain of tactics is a very difficult task. If the correspondence is of the one-one type, then the tactic is the exact counterpart of the process, so that the observation of the tactic gives exactly the same information that would be given by the observation of the process. In most situations this is not probably the case.

A more reasonable assumption is that the ocrrespondence is of the many-one type, so that each process corresponds to one tactic though several processes may converge into the same tactic. In terms of our observations this would correspond to the situation in which the same observable behavior may result from different processes. In all likelihood this type of correspondence prevails in most situations.

But if the correspondence is of the one-many type, then each process may originate entirely different tactics, leading to different solutions, and confusion would prevail. From these assumptions it follows that the experimental technique that we use to evaluate processes requires a one-one or a many-one type of correspondence between processes as the domain and tactics as the codomain.

As previously stated, subjects are "within limits" free to ask questions. The major difficulty of this assumption lies in the parenthetical expression "within limits". The previous research cited lends support to the validity of this assumption. It is important to note that once a given question has been asked, the likelihood of asking or not asking one of the remaining questions will change. This will greatly depend on each subject, on the questions that he has asked previously, on how he perceives the goal, etc.

Another assumption is that tactics do reflect in some manner the information that the subject searches for, how he evaluates this information, how thoroughly he explores sub-goals and hunches, etc. This is a rather universal assumption in psychological research in the sense that observable behavior corresponds to covert behavior in some specified manner.

The same general considerations that were previously made to explain the correspondence between processes and tactics may be used to express the correspondence between specific events in the process and specific observations in the tactic. Notice that in the case of the many-one relationship each event in the process would correspond to one and only one event in the tactic, though several events in the process may converge into one observable event in the tactic.

Lastly it seems much more fruitful to study the tactic rather than the final answer. The tactic by definition is more complex and prior to the final answer. This should allow for greater differentiation between performances.

The problem solving technique as used in this study

The specifications pertaining to the problems used in this study resulted from a variety of theoretical considerations and experimental results obtained in several separate studies. The formulation of the problems, the scoring procedures used, and the areas explored have gone through several changes. Meanwhile the hypotheses tested and the theoretical considerations have come into sharper focus. This has allowed us to develop a plan for research in which the contribution played by the formal properties of the problem and their manner of presentation can be experimentally isolated.

Before describing the main features of the technique used in this study, a brief historical sketch of its development will be outlined. Throughout this developmental history, the following points will be considered the characterization of (a) cognitive processes in a problem solving situation by the questions that a subject asks (tactic), (b) the development of a scoring procedure to characterize the performance of a subject independent of a specific sample, (c) the more precise identification of what is "built in" the problems, that is, a more accurate control of the instruments used, (d) how these instruments differentiate between individuals, and (e) how the logical structure of a problem and the manner of expression used (language) affect a subject's tactic -- and by influence, his processes -- so that cognitive styles may be described.

Originally problems were built according to specific content areas like medical diagnostic skills (Rimoldi, Haley, Fogliatto, 1962), process of Rorschach evaluation (Tabor, 1959), psychotherapy processes (Gunn, 1961), the Problem Solving and Information Apparatus (PSI) (John, Rimoldi, 1955), Training in Problem Solving (Rimoldi, Devane, 1961), mathematical problems (Rimoldi, 1967), etc. In essence we were experimenting with (a) the contents of each problem, (b) which manner of presentation was more appropriate for each specific content area, for instance in the Test of Medical Diagnostic Skills careful consideration was given to whether the information should be given by writing, actual x-ray films, or photographs, etc., (c) evaluation of tactics.

In the early years we analyzed independently the number of questions asked, the type of questions, and their order. The number of questions asked is a simple score to obtain though its meaning is rather complex. It was demonstrated that this score differentiated clearly between levels of training (Rimoldi, Haley, Fogliatto, 1962 ; Rimoldi, Devane, 1961 .; Rimoldi Fogliatto, Haley, Reyes, Erdmann, Zacharia, 1962 ; Rimoldi, Fogliatto, Erdmann, Donnelly, 1964). It is still kept as an integral part of our scoring system.

The utility index of each question was defined as the proportion of subjects that chose a given question in a given sample, the assumption being that if a question was perceived by the members of a group as more "useful" than others, then it would be selected more often.

It was possible to show that different questions had different utility indexes and that the same question may have different utility indexes depending on the educational level or other characteristics of the subjects in the sample being tested (Rimoldi, Haley, Fogliatto, 1962 ; Rimoldi, Devane, 1961). Significant differences were related to expected levels of medical training, levels of education, etc.

For a given sample, questions can be rank ordered in terms of the decreasing value of their utility indexes. Accumulating these rank ordered utility indexes and plotting them against successive steps in the sequence, a maximum curve is obtained. This corresponds to the ideal performance that would represent a tactic in which the association is maximal between utility index and order in the sequence. The performance curve corresponding to a specific tactic can be drawn on the same graph and compared with this maximum curve (Rimoldi, Devane, Haley, 1961). Plateaux would occur whenever questions with a utility index of zero are chosen. The number and length of plateauz occurring in a given tactic was shown to be a significant component in differentiating problems and samples (Rimoldi, Fogliatto, Haley, Reyes, Erdmann, Zacharia, 1962 ; Rimoldi, Fogliatto, Erdmann, Donnelly, 1964; Haley, 1960). On the basis of these utility indexes, it was possible to differentiate the difficulty of problems (different problems, same sample of subjects) and level of samples (same problem, different samples)(Rimoldi, Haley, Fogliatto, 1962 .; Rimoldi, Devane, 1961 .; Rimoldi, Fogliatto, Haley, Reyes, Erdmann, Zacharia, 1962 ; Rimoldi, Fogliatto, Erdmann, Donnelly, 1964). But useful as they might be, utility indexes, strongly depend on sample norms and do not consider the order in which a question is asked, and thus they are weak indexes of the process. Still the remarkable association shown between this rough method of evaluation of performances and several independent criteria was a strong indication of the "validity" of the technique as a whole.

In order to introduce the variable "order in the sequence" we explored the preferred location of a given question in a collection of tactics. It was found that some questions were preferrentially selected at the beginning of the process, others at middle or at the end of it, and still others almost anywhere in the sequence. It was also found that this preferred location might vary with the sophistication of the subjects. The relationship between the utility index of a question and its preferred location in the tactic was analyzed by coefficients as well as Kendall's W (coefficient of concordance; Rimoldi, Haley, Fogliatto, 1962 ; Rimoldi, 1955).

The method of pattern analysis was developed to investigate this order component (Rimoldi, Grib, 1960). But again, none of these procedures provided a "single" evaluation of the tactic. Nevertheless we were becoming progressively aware that there were other important variables related not only to the subject but more specifically to what "we put" in the problem.

The next step towards a scoring system was to determine the frequency with which each question was selected in each possible order. This can be represented by tables in which columns represent questions and the row, order in the tactic. The entries in the cells correspond to frequency of selection. Since a question may be asked only once, performance can be represented using probabilities as defined in random sampling without replacement (Feller, 1957). This means that if at a given step anyone of the possible questions is equally likely to occur, then at step k its probability of seclection will be $\frac{1}{n-k+1}$, where n is the total number of possible questions presented with the problem and k is order ($1 \leq k \leq n$) so at the first step we will obtain $\frac{1}{n}$, and when k = n we obtain $\overline{1}$, that is the last questions asked is completely determined. These values could then be used to test for independence between questions and order. This is an important consideration since we are assuming that the order of occurrence of a question depends on what has preceeded it. In other words in the tables just described, covariance between questions and order plays an important role, and any statement based on scores obtained by using this information should consider the association between questions and order. Nevertheless these values were used as an approximation to score individual tactics, by accumulating the values corresponding to the questions asked in the order in which they are asked. This allowed the description of individual performance curves, including their length and the existence or non-existence of plateaux. We call this approach scoring using group norms. Clearly the procedure is heavily dependent on the sample used to define these values and violates some of the purposes that we initiated previously.

On the other hand if all the members of a group agree absolutely on the questions they ask and when they ask them, then the corresponding tabular representation could be given by a table in which in each row and in each column there will be (n - 1) zero entries (for a problem with n questions) and only one filled in cell. That is, each question is asked in one and only one order by all the members of the group. This represents maximum agreement among subjects. The "chance agreement" would correspond to cells filled in using the $\frac{1}{n-k+1}$ formulation. For descriptive purposes an index of agreement can be defined using the transformation of information theory (Attneave, 1959) so that maximum agreement would correspond to minimum uncertainty values. Then the performance of any sample of subjects in a given problem can be readily compared in terms of its departure from this minimum uncertainty or vice versa in terms of its distance from the uncertainty value obtained under random sampling without replacement. These evaluations of agreement are strictly speaking, description of sample performances and do not say that a sequence that gives maximum agreement is necessarily the correct one. For instance assume questions a, b and c. Any of the sequences: a, b, c; a, c, b; b, c, a; b, a, c; c, a, b; and c, b, a, will give minimum uncertainty value if all the subjects follow this sequence exactly but only one of them may be correct. This approach was extensively used in several studies (Rimoldi, Fogliatto, Haley, Reyes, Erdmann, Zacharia, 1962 ; Rimoldi, Fogliatto, Donnelly, 1964; Rimoldi, Aghi, Burger, 1968 ; Vander Woude, 1969; Rimoldi, 1967) and in spite of its shortcomings can be used to establish a relationship between agreement in group performance and its logic of approach.

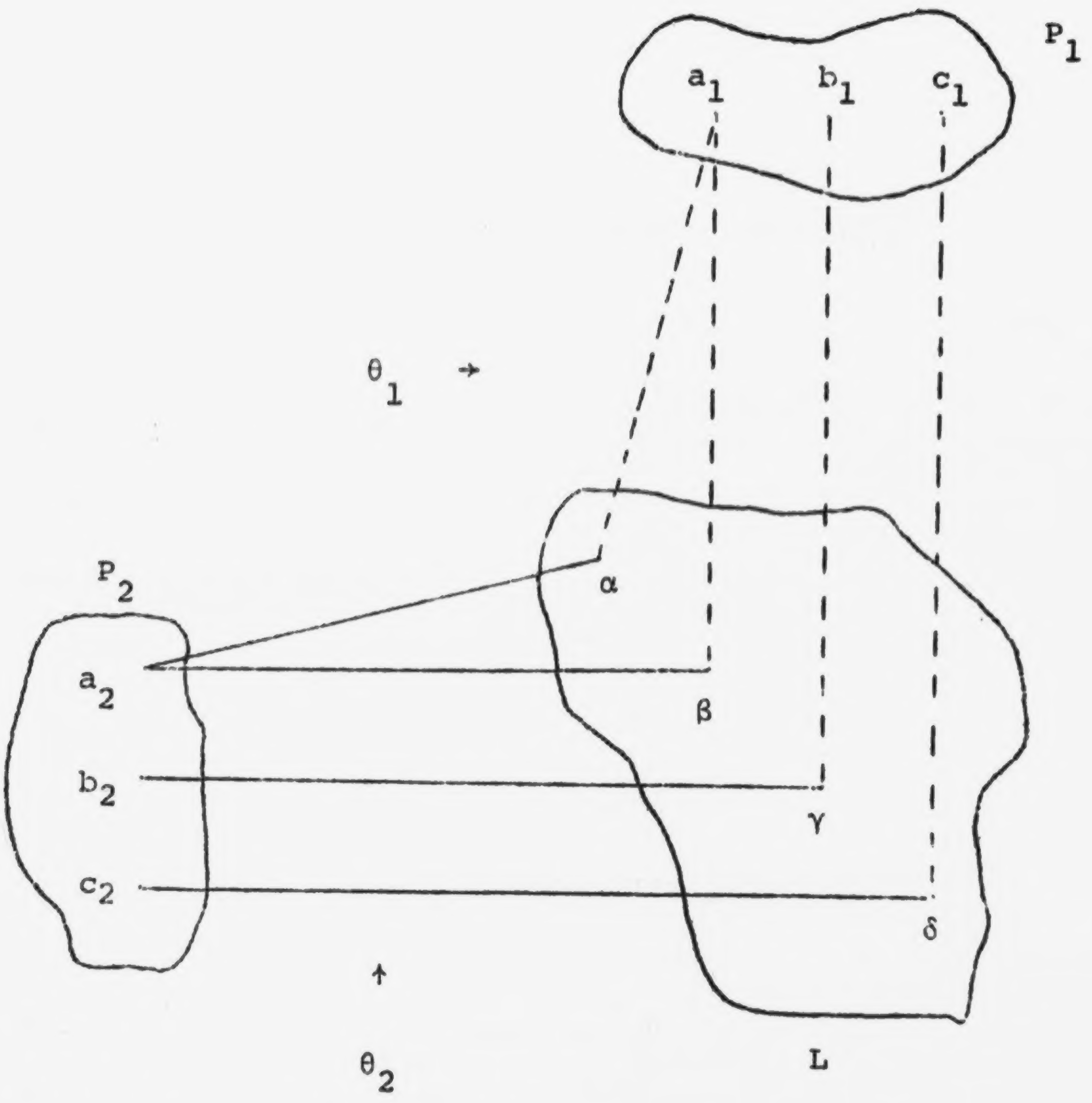
In studying how subjects solved certain mathematical problems, it was observed that rather complex logical relationships, like those in Newton's binomial or in the Pythagora's theorem could be dealt with if these structures were presented in a non-symbolic, everyday language. Of course, the children examined in that research did not know that in order to solve the given problem they were operating with the relational systems inherent to some complex logical formulation. As is often the case, the question is whether subjects can operate with a complex structure other than if they are aware that they are using logic or mathematics.

It became clear that if the structure of the problem was known then a scoring procedure based on this structure could be developed, that that this scoring procedure was independent of sample norms, though it could be used "a posteriori" to characterize samples. It was hypothesized that the same structure could be presented in different manners, so that a given performance could be understood as a function of both the structure of the problem (intrinsic difficulty) and the manner of presentation or language used (extrinsic difficulty) (Rimoldi, Haley, Fogliatto, Erdmann, 1963). Our most recent work has been based on the possibility of separating the logical structure of a problem from the language used. The problems used in this research were devised using these two concepts.

Logical structure and language in the schema pulling out method of scoring tactics

These two concepts have been extensively used in the study of cognitive processes but often it is difficult to say whether an experimenter is referring to one of them, to both or to neither of them. The logical structure of a problem can be defined by the system of relations that hold between certain elements in a defined set. By language is meant the manner in which the logical structure is presented provided a correspondence has been defined between the elements of the language and those of the logical structure. The language may consist of words, abstract symbols, objects, etc. Since structure can be presented in several languages, these two concepts are in most cases operationally independent.

Assume that the logical structure L (Figure 1) consists of element α, β, γ and δ, with specified relations among themselves. The images of these elements can be obtained under different θ_i mappings. Each mapping represents a language so that $\theta_i\ (L) = P_i$, where P_i is the problem built using language i. The correspondence between elements in the logical structure and elements in the language can be specified. As indicated in Figure 1 the element a_1 in problem P_1 corresponds to a many-one relationship so that a_1 is the image of both α and β in L when language θ_1 is used. For another language (θ_2), a_2 also corresponds to α and β, but the mapping of the logical structure is performed using language θ_2. This generates problem P_2. The element γ has for image b_1 under mappings θ_1 and b_2 under mapping θ_2, and similarly for c_1 and c_2. Under these conditions, problems P_1 and P_2 are isomorphic so that $a_1 \leftrightarrow a_2$, $b_1 \leftrightarrow b_2$ and $c_1 \leftrightarrow c_2$. Most of the problems used in this study meet these specifications. In this sense a problem can be defined as a mapping of a logical structure.


P_1
a_1
b_1
c_1
θ_1 →
P_2
α
a_2
β
b_2
γ
c_2
δ
↑
θ_2
L

$\theta_1 (L) \rightarrow P_1$

$\theta_2 (L) \rightarrow P_2$

Figure 1

Correspondence between Isomorphic Problems

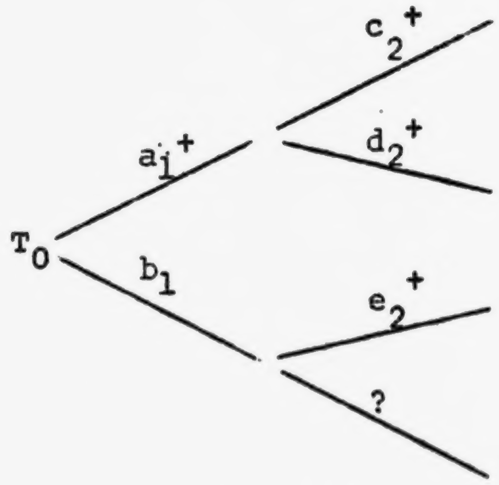

c_2^+
a_1^+
d_2^+
T_0
b_1
e_2^+
?

Figure 2a. Structure 31

$a_1 \leq T_0$, $b_1 \leq T_0$, $c_2 \leq a_1$,

$d_2 \leq a_1$, $e_2 \leq b_1$, $? \leq b_1$

Ideal Tactic: $a_1 \rightarrow e_2$

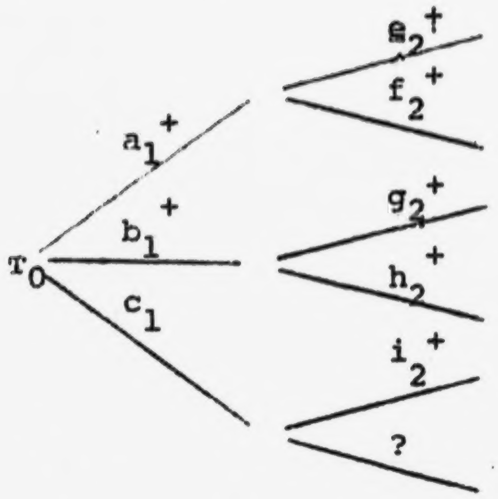

e_2^+
f_2^+
a_1^+
g_2^+
b_1^+
T_0
h_2^+
c_1
i_2^+
?

Figure 2b. Structure 33

$a_1 \leq T_0$, $b_1 \leq T_0$, $c_1 \leq T_0$,

$e_2 \leq a_1$, $f_2 \leq a_1$, $g_2 \leq b_1$,

$h_2 \leq b_1$, $i_2 \leq c_1$, $? \leq c_1$

Ideal Tactic: $a_1 \leftrightarrow b_1 \rightarrow i_2$

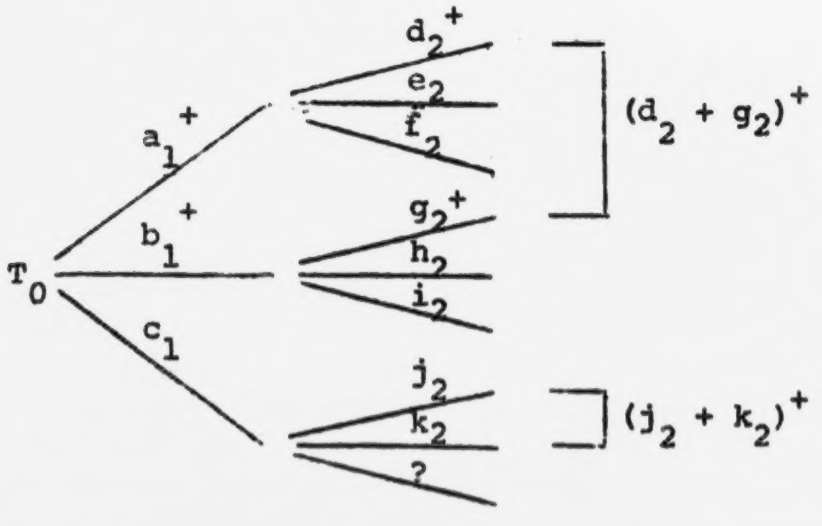

d_2^+
e_2
f_2
$(d_2 + g_2)^+$
a_1^+
g_2^+
b_1^+
h_2
T_0
i_2
c_1
j_2
k_2
$(j_2 + k_2)^+$
?

Figure 2c. Structure 35

$a_1 \leq T_0$, $b_1 \leq T_0$,

$c_1 \leq T_0$, $d_2 \leq a_1$,

$e_2 \leq a_1$, $f_2 \leq a_1$,

$g_2 \leq b_1$, $h_2 \leq b_1$,

$i_2 \leq b_1$, $g_2 \leq c_1$,

$k_2 \leq c_1$, $? \leq c_1$

Ideal Tactic: $a_1 \leftrightarrow b_1 \rightarrow (j_2 + k_2)$

In a restricted sense, a thinking process can be understood as an attempt to make explicit and communicable to oneself or to others the formal properties of a problem (Rimoldi, 1967). The tactic indicates how a given subject attempts to make explicit the logical structure of the problem. Depending on the subject, the logical structure and the language used, different tactics may be used. Clearly if the language used is not understood by the subject, then he cannot solve the problem except by chance, and this can be observed in the tactic that he follows. Contrarywise, if a logical structure is made explicit and communicable in at least one language, then it can be said that the subject can operate with the formal structure involved in the problem.

As the section on Method will indicate, several structures were used in this research. Each one of them gave rise to several isomorphic problems presented in different languages. In that section several properties of the scoring methods are presented and in Figure 2 the structures are presented. The more specific approach used for scoring these problems will be discussed in the section on Method. Scoring is a function of the number of questions, their type and their order and it is independent of any sampling evaluation. Indeed, the subject following the ideal tactic will obtain the maximum possible score. Ideal tactics are those in which the subject accumulates all the information needed to solve the problem, without redundancies, without order reversals, without irrelevant questions, and in the most parsimonious manner. Whenever these requirements are violated, his score will be less than that corresponding to the ideal tactic.

It is thus clear that the scoring procedure is independent of any sample evaluation. As a matter of fact the score that a subject obtains is exclusively the result of the structure of the instrument used and of his ability to solve the problem. Clearly subjects of different ability levels or with different personality characteristics, etc., will perform differently if the problems are sensitive to these variables. The method used to score tactics that has been called schema pulling out method implies the following: 1) we examine each tactic for redundancies, order reversals, etc., after having eliminated from such a tactic questions that are irrelevant. By irrelevant we mean those that do not contribute information pertaining to the solution of the problem. After this has been done, a value is assigned to each one of the relevant questions, these values are combined and they are weighted by the total length of the tactic that the subject followed, including relevant and irrelevant questions as well. This is why the method has been called schema pulling out method. Irrelevant questions are first pulled out from the tactic and then are reintroduced so that the final results are weighted by the total number of questions asked.

In several studies in which this approach was followed, it was possible to observe that a subject's tactic depends on both the structure of the problem and the language used. A factor analysis of problems (Rimoldi, 1970) indicated that factors corresponding to language and factors corresponding to logical structures could be clearly separated. Further, it has been shown previously (Rimoldi, Aghi, Burger, 1968) that if the structure of the problem is simple, then there is no interaction between language and age (from 7 to 13 years of age) but for more complex structures this is not the case. It was shown that at a certain age level certain new languages take precedence over languages already well developed. Also (Vander Woude, 1969) it was shown that prelingually deaf children are able to solve relatively complex problems provided they are presented in a language that they can understand. Finally the effects of language on performance in isomorphic problems was shown for the ages 9 to 78. In all cases the problems presented in an abstract-symbolic language showed lower scores than problems presented in ordinary verbal language (Rimoldi, Vander Woude, 1969).

General theoretical and methodological considerations

The correspondence between processes and tactics has been analyzed in previous pages. Among these we have defined the ideal tactic that follows closely the formal structure of the problem as given. If isomorphic problems (same structure, different languages) are administered, then the contribution of the language used (extrinsic difficulty) and of the logical structure (intrinsic difficulty) can be separated.

If the subject cannot understand the logical structure built in a problem, then no matter what language is used, his tactic will be widely separated from the ideal tactic. But if he does understand the structure, then there will be a language in which he can express it. In the extreme case, if there is not a language and if he has enough ingenuity, he might be able to invent a language in order to express the complexity of the structures involved.

In terms of these concepts, it seems risky to conclude that a subject cannot understand a certain logical structure because he cannot solve a problem that implies such a structure. The reason may be simply that this subject does not understand the language used.

With certain languages it is possible to express a greater variety of logical structures than with other languages (Rimoldi, 1970; Rimoldi, 1967; Rimoldi, Aghi, Burger, 1968; Vander Woude, 1969; Rimoldi, 1969). This was observed in the factor analytic results reported above. Languages by means of which a great number of logical structures can be expressed will be called onto languages. In this sense some languages approximate this onto property more than others. The usual spoken and written languages are of this type. Nevertheless, two languages may not be isomorphic and their degree of "ontoness" may vary. For instance it is well known how difficult it may be to translate very precisely concepts from one spoken language into another. It seems also intuitively clear that with signs, gestures, etc., only certain structures can be expressed and not always with great precesion. In some cases it might be necessary to develop new symbols (for instance in mathematics and in sciences in general) to express some structures more efficiently and without ambiguity. But these languages may not necessarily be appropriate to express other structures.

This implies that besides their completeness (ontoness) languages may vary in precision. One-one languages are by definition of complete precision, so that each element in the language represents one and only one element in the structure and vice versa. Incidentally, since one-one relationships imply the existence of an inverse, then with this particular type of language, whatever the subject does is the exact counterpart of whatever the subject thinks. This precision may be reached at the expense of reducing the number of logical structures that can be expressed. That is, certain structures can be better expressed using one specific language, but this specific language might be inappropriate to express other structures.

Methods

Sample

The sample* for this study consisted of 705 elementary school boys between the ages of 7 and 12 from a large metropolitan center in the midwest of the U.S.A. They attended seven different Catholic schools, were predominantly of Polish descent, and lived in middle class neighborhoods. The selection of the subjects was made on the basis of responses received from the parents. Letters explaining the purpose of the study were sent to the parents of all the boys of the above mentioned age. Parents' permission to test their boys was requested. Only those (boys) whose parents answered affirmatively were used for the study.

As indicated in Table 1, 705 subjects were tested during the first year, 396 of these the second year and finally of these 396, 352 were tested the third year. The dotted line in the table separates 11 and 12 year olds in the first year of testing from the earlier ages. The reason being that 11 and 12 year olds would be 13 and 14 years old respectively in the second and third year of testing and these ages were not contemplated in the design of the study. However 11 and 12 year olds were tested in the first year in order to establish a base line for further comparisons. The characteristic of the design partly explains the large difference in the total number of subjects in the second year of testing compared with that of the first year (from 705 to 396 of which 256 were 11 and 12 years old in the first year of testing). Looking at Table 1 horizontally, it is seen that the original sample for each age is reduced approximately 10% in the successive years. This was due to reasons beyond our control such as subjects changing schools, moving out of the city and absenteeism at the time of testing.

For every subject the following data was obtained from the school records: (1) Age from the nearest birthday to the time of the first year testing, (2) I.Q. All the schools administered Otis Intelligence Test as a routine procedure. The results of the latest testing were used to classify the students into high, medium and low using the criterion to be described later in the study. (3) Grades. Since all the schools offered to same subjects and had the same criterion for evaluation of student performance, the grade average for the first year of testing was obtained. The

---

*We are indebted to Dr. Wozniak, Dean, School of Education, Loyola University and Sister Stephenette, C.S.S.F., Educational Consultant for the Schools directed by The Felician Sisters in Illinois, Lecturer, School of Education, Loyola University, for their indispensable and amicable cooperation in providing us the sample for this study.

	First Year	Second Year	Third Year
	7 (104)	8 (98)	9 (88)
	8 (105)	9 (94)	10 (85)
	9 (116)	10 (88)	11 (82)
	10 (124)	11 (110)	12 (94)
	11 (129)	12 (5)	13 (3)
	12 (127)	13 (0)	14 (0)
Total	705	396	352

Table 1

Description of the sample. In each line the top number refers to age and the number in parenthesis (lower row) to number of subjects.

academic courses on which the grade average is based were: Reading, English, Spelling, Handwriting, Arithmetic, History, Geography, Science, Christian Doctrine, Music and Art. (4) SAT scores (Standard Achievement Test). (5) Father's occupation: the following six categories were employed:

0 - Deceased father
1 - Unemployed
2 - Laborer, cook, etc.
3 - Office worker, factory supervisor, etc.
4 - Management, own business, etc.
5 - Professional

(6) Home conditions -- that is, whether the child lived with both parents or comparable guardians:

0 - Separated or divorced
1 - Family intact

These six items of information were coded together with the results of the testing for each student identified by school and number. All this information was put on IBM cards used for the statistical analyses.

Problems

The 27 problems basic to the analysis are presented in Table 2. For each of the structures 31, 33 and 35, there are 3 languages: Picture, Verbal A and Verbal B. For each of the 3 testing years, the problems are different except for those bracketed, that is, $31VB_1 = 31VB_2$, $33VB_1 = 33VB_2$, and $35VB_1 = 35VB_2$. Problems of a given structure are isomorphic except for the few exceptions mentioned below. A copy of each problem may be found in Appendix A.

Each problem consisted of 10 elements or questions, possibilities for which are illustrated as branches marked "+" designating available elements in a problem, along with the statement of the problem T_0. The remainder of the 10 elements are irrelevant questions.

Problems of the 31 series are represented in Figure 2. All nine of these problems are isomorphic. The elements marked "+" remain constant for each problem. Any combination of these marked questions constitutes a tactic, an attempt at solution. $a_1 \to e_2$ represents the ideal tactic; $c_2 \leftrightarrow d_2 \to e_2$ an acceptable one, $b_1 \to c_1$ an impossible one for b_1 is not available (no "+").

Of the 33 series, all but $33P_2$ and $33VA_2$ are isomorphic and fit the pattern in Figure 2. In these two cases at the risk of

	Structure		
	31	33	35
1st Year	$31P_1$ $31VA_1$ [$31VB_1$]	$33P_1$ $33VA_1$ [$33VB_1$]	$35P_1$ $35VA_1$ [$35VB_1$]
2nd Year	$31P_2$ $31VA_2$ [$31VB_2$]	$33P_2$ $33VA_2$ [$33VB_2$]	$35P_2$ $35VA_2$ [$35VB_2$]
3rd Year	$31P_3$ $31VA_3$ $31VB_3$	$33P_3$ $33VA_3$ $33VB_3$	$35P_3$ $35VA_3$ $35VB_3$

Table 2

Problems used in the study classified in terms of structure, language and year of administration.

not having strictly parallel problems across the years, an effort was made to assess the sensitivity of the procedure to slight changes in the presentation of the same structure. In $33P_2$ an additional element $(f_2 + h_2)^+$ appears in the problem and also consequently there is one less irrelevant question. In $33VA_2$ two elements f_2 and g_2 were deleted and two irrelevant questions added. Despite these small variations, the ideal tactic is the same for all 33 series problems, that is $a_1 \leftrightarrow b_1 \rightarrow i_2$.

In the 35 series, problems $35P_1$, $35VA_1$, $35VA_3$, $35VB_1$, $35VB_2$, and $35VB_3$ are isomorphic and represented in Figure 2. Notice that both d_2^+ and $(d_2 + g_2)^+$ are given as well as $(j_2 + k_2)^+$. These problems have four irrelevant questions. The three remaining problems of this series have the same structure though some of the given elements are different. $35P_2$ includes $(d_2 + k_2)^+$ instead of $(d_2 + g_2)$ and k_2^+ additionally (only three irrelevant questions). $35VA_2$ uses f_2^+, h_2^+, and j_2^+ instead of d_2, i_2, and $(d_2 + g_2)$. In $35P_3$, e_2^+ and $(h_2 + k_2)^+$ are substituted for d_2 and $(d_2 + g_2)$. The ideal tactic for all 35 series problems is $a_1 \leftrightarrow b_1 \rightarrow (j_2 + k_2)$.

The following eight problems were also used but not included in the analyses to date. Three seriation problems which involved simple orderings and three redundant picture problems were given the first and second years of testing. Redundant picture problems were parallel to the picture problems above except for the addition of a question that contained all the information presented in T_0, the statement of the problems. Problems 42 and 31A, not original problems for this study, were employed as a base line to compare performance levels with previous researches.

All these 35 problems are presented in full in Appendix A. Some additional problems similar to the ones above were also utilized as training devices but are not here described.

Testing

As indicated in the proposal, one of the main objectives of this research is to determine at what age and ability level children are able to solve problems of a specified logical structure with different modes of presentation (languages). We attempted to determine the lowest age at which children are able to operate with problems of known logical structure and mode of presentation. In a way this could be interpreted as trying to establish "thresholds" for the two variables, structure and language. There were strong indications from our previous research (Rimoldi, Aghi, Burger, 1968; Rimoldi, Vander Woude, 1968; Rimoldi, 1967) that at certain ages for problems with a given logical structure, the mode of presentation that is most effective changes markedly. This is borne out by the fact that while at early ages problems presented with a specific language give satisfactory results; at a more advanced age the same problem is less satisfactory and sometimes inferior to another mode of presentation.

It was hypothesized that the difficulty increased with the logical complexity of the problems, that is, 33 is more difficult than 31 and 35 is more difficult than 33. Also is was hypothesized that the picture problems were easier than the verbal ones and the verbal in turn easier than the symbolic ones. Further it was assumed that the seriation problems were the easiest.

Previous research (Rimoldi, Aghi, Burger, 1968) indicated that the length of the testing session should be about 1-1/2 hours long to gain and maintain the interest of the subjects. This time limit restricted the administration of all the problems to each subject. All these matters were considered in the testing design which is described below for each year.

First year testing

a) Seven and eight year old group:

Each subject was tested individually and was given a minimum of nine problems. The average session was 1 to 1-1/4 hours long. The subjects were first given the three seriation problems followed by a practice picture problem. After this the picture problems $31P_1$, $33P_1$ and $35P_1$ were administered in that order, followed by the redundant picture problems $31PR_1$, $33PR_1$ and $35PR_1$. This was done for all subjects regardless of whether they were able to reach a successful solution. At this time if the subject performed well, the tester continued with the verbal problems, otherwise the testing session terminated. By "performing well" is meant that the subject on the whole followed either the ideal tactic or any of the good tactics (these terms will be explained in detail in the following section on scoring). If it was decided to continue with the verbal problems, they were administered in the following order: 31A, $31A_1$, $33A_1$, $35A_1$, $31B_1$, $33B_1$, $35B_1$ and 42. If at any time after the first two problems, the subject could not perform well in one of the problems, the testing was terminated.

b) Nine year old age group:

Each subject was tested individually in sessions of about 1 to 1-1/4 hours long. The subjects were first given a practice verbal problem followed by 31A, $31A_1$, $33A_1$, $35A_1$, $31B_1$, $33B_1$, $35B_1$ and 42 in that order. If it was clear to the tester after administering two problems, that the subject was failing to perform well, he was not given the remaining problems. Instead he was given the picture problems, beginning with practice picture problems and followed by $31P_1$, $33P_1$, $35P_1$, $31PR_1$, $33PR_1$, $35PR_1$ in that order. If he still failed to perform well, the seriation problems were administered.

c) Ten, eleven and twelve year old group:

The order in which the problems were administered was the same as for age nine. The testing was performed, however, in groups of 10-20 instead of individually. If a subject's protocol indicated that he was not performing well, he was retested individually with the picture problems. The group sessions lasted about 1-1/2 hours.

In summary, the thresholds for problem solving were established for testing all the subjects within a predetermined interval of hypothetical problem difficulty and then proceeding to a higher level if the subject was performing well or to a lower level if his performance was poor.

Second year testing

a) Seven* year old group:

Again each of these subjects was tested individually and was given a minimum of nine problems. The average session was 1 to 1-1/4 hours long as in the first year of testing. The order in which the problems were administered was also the same: first the three seriation problems followed by three picture problems, $31P_2$, $33P_2$ and $35P_2$. Then the three redundant picture problems $31PR_2$, $33PR_2$ and $35PR_2$ were given in that order. After this was done, the tester decided whether or not to continue with the verbal problems depending on the performance of the subjects. If it was decided to continue with the verbal problems, they were administered in the following order: first the practice verbal problem, then $31A_2$, $33A_2$, $35A_2$, $31B_2$, $33B_2$, $35B_2$ and 42 in that order. Again the testing session terminated as soon as it was evident to the tester that the subject was not performing well.

b) Eight year old group:

Each subject was tested individually as during the first year of testing and the procedure and order of administration was identical to that during the first year. Thus the subjects

*For the sake of clarity, age groups are always here identified by their first year age, for example here seven year olds are actually eight and also eight year olds would be ten in the third year of testing, etc.

were first given a practice verbal problem, followed by problems $31A_2$, $33A_2$, $35A_2$, $31B_2$, $33B_2$, $35B_2$ and 42 in that order.

c) Nine and ten year old age group:
The order in which the problems were administered was the same as in the previous year. This year the nine year olds like the ten year olds were tested in groups of 10-20 subjects. Again if the subject's protocol indicated a poor performance, he was retested individually with the picture problems.

Thus the problem solving thresholds were established for the second year also.

Third year testing

a) Seven year old group:
Each subject was tested individually in sessions of about 1 to 1-1/4 hours. The subjects were first given a practice verbal problem followed by problems $31A_3$, $33A_3$, $35A_3$, $31B_3$, $33B_3$, $35B_3$ and 42 in that order. The procedure for determining whether to continue or to terminate the testing session was the same as during the first and second year of testing.

b) Eight and nine year old group:
The order of problems was the same as for the seven year old group. The testing was again performed in groups of 10-20 subjects. The group session lasted for about 1-1/2 to 2 hours as during the previous two years of testing.

Testers:

The testing was accomplished by ten Loyola Psychometric Laboratory male and female personnel who were at the level of advanced graduate students. Efforts were made to control individual-tester biases. All the testers were instructed and trained in live situations by an experienced member of the laboratory. In addition, the testers were generally assigned in a random fashion to individual subjects or groups of subjects of various ages over each of the three years of testing. The tester's function was to give preliminary instructions, explain a sample problem, and present the test problems. If further instructions were necessary once testing commenced, deliberate efforts were made to remain neutral, that is, neither to facilitate or hinder the subject's performance.

Scoring procedure

The method used here is a refinement and a redefinition that resulted from the comparison of twenty different scoring methods in a Master's dissertation, by Mr. T. Chlapecka.

The purpose of this analysis was to design a scoring procedure that would be sensitive to the approach of the subject to the logical structure of the problem. The second and more difficult requisite was that this measure should have a significant degree of invariance to the superficial peculiarities of a problem, that is the number of questions, the ratio of relevant to irrelevant questions, etc. This means for example that a chance performance should have the same score in any problem. Here that score is set to **zero**, the score of the ideal tactic is set to 1.00, and the minimum score is -1.00. A score $Si = .35$ for example should reflect as accurately as possible the same degree of problem solving proficiency regardless of the problem, and likewise for other values $-1.00 \leq Si \leq 1.00$.

The approach followed considers how each tactic approximates the logical structure of the problem. By tactic we mean the sequence of questions that the subject asks. Approximating the logical structure involves asking the more general questions first and thereafter questions of increasing specificity. The ideal tactic fulfills these requirements, that is: maximum correspondence between the generality of the question and its position in the tactic with the minimum number of questions that exhaust the information necessary to solve the problem. In the scoring system used, these ideal tactics will obtain the maximum score. Scores are lower to the degree that they violate the above conditions, that is: reversals in order, irrelevancy and/or redundancy, lack of parsimony, failure to choose meaningful questions, etc.

The questions that the subject may ask can be classified into two major categories: relevant questions are those that provide information that is pertinent, irrelevant questions are those that provide no pertinent information. The relevant questions in turn can be further classified into subclasses defined by their degree of generality. With reference to Figure 2 (a, b, c) in which the branches exemplify questions, each successive branch of the tree represents more specific questions. The degree of specificity is directly related to the magnitude of the subindexes. An order reversal will occur when a more specific question is asked prior to a more general question. A group of specific questions is equivalent to a more general question where the information embodied in them is the same as the information provided by the more general question. If a subject asks both a general question and some or all of the equivalent specific questions, those specific questions are considered redundant independently of their position relative to the general question.

Assigning a score to an observed tactic involves several stages:

1) Like in the "schema pulling out method" (CRP 1449 and CRP 2199), all irrelevant and redundant questions are set aside from the observed tactic. This procedure reduces an observed tactic to one of the possible basic tactics.

2) The elements remaining in the basic tactic are then analyzed for order reversals. For instance, reference to Figure 2c indicates that problems built around structure 35 have two types of questions: a_1, b_1, and c_1 of maximum generality and d_2, e_2......k_2 of less generality or greater specificity.

In the scoring system reversals of questions within each order of generality are not considered. Thus, for the case of Figure 2c, the sequences a_1, b_1, c_1, and c_1, b_1, a_1 are identical and so are d_2, f_2, h_2, k_2 and f_2, h_2, k_2, d_2 and so forth. But f_2, b_1, c_1 implies reversals since the order of their occurrence in terms of the specifications set forth previously should be b_1, c_1, f_2 or c_1, b_1, f_2. The number of more general questions that appear in the tactic determines the number of positions in which the less general questions may occur. That is, questions b_1 and c_1 determine three possible positions for any specific question: either before, in between, or following them. If a specific question follows a general question, it is arbitrarily assigned a positional number of 1. Other positions are assigned values related to the number of steps that they are removed from the "logical" order. So question f_2 has the following values: 1 if in sequences c_1, b_1, f_2 or b_1, c_1, f_2, 2/3 if in sequences b_1, f_2, c_1, or c_1, f_2, b_1, and 1/3 if in sequences f_2, b_1, c_1 or f_2, c_1, b_1.

The general formula to determine these positional numbers is: $a_{pj} = 1 - \frac{j}{k}$, where a_{pj} is positional number for a question p in position j, j corresponds to the number of steps that the question p is removed form its "logical" order, and k is number of possible steps. This formulation can be extended to problems with any values of k, where $0 \leq j \leq (k-1)$.

3) With the positional numbers as defined in 2) above, a matrix L is built in which the rows correspond to all the questions presented with a problem and the columns to the basic tactics as specified in 1) above. In the cells of matrix L, the corresponding positional numbers are entered, the values for the irrelevant and redundant questions being zero. An example of such matrix is given later.

4) Each question in the problem is assigned a value in terms of the information it provides. Two conditions determine the information values (I) assigned to each question. First, assigning a score of 1.00 to the ideal tactic limits the information values of the x relevant questions. The more relevant questions in a problem, the smaller will be the information weight of each. The second condition relates the information weight of a general question to those of its equivalent class. The information weight of a question at level t is defined as:

$$I_t = (r + \frac{r-1}{r})\ I_{(t+1)}$$

where r is the number of elements in the equivalent class (sub-branches). For example in Figure 2a a_1 has two sub-branches c_2 and d_2 (r=2), and $I_1 = (2.5)\ I_2$.

A (1 x n) row vector W (W_1, W_2.....W_n) may be used to represent these information values for a given logical structure. A sample follows. Notice the information value of a given general question is greater than the sum of information values of its sub-branches by an amount directly related to the number of branches r. This process allows a weighting for parsimony or economy of a tactic. Consider in Figure 2a the two tactics a_1, e_2 and b_2, c_2, e_2. Both exhaust the information necessary to solve the problem, yet a_1, e_2 is more efficient and receives a higher score. A detailed rationale for this procedure may be found in the above mentioned thesis by T. Chlapecka.

5) The score of basic tactic (x_i), where i = 1.....m, can be represented as a linear function of the information weights and of the positional numbers so that, $x_i = w_1 a_{1i} + w_2 a_{2i} + \ldots\ldots + w_n a_{ni}$. Similarly, for all the other basic tactics.

In the $(n \times m)$ matrix L, all the basic tactics corresponding to a given structure may be represented by the (m) columns, while the (n) rows correspond to all the questions in the problem. The $(1 \times n)$ row vector X, gives the score for all the basic tactics and is obtained by the matrix multiplication:

$$X = WL$$

In Figure 2a for problem $31P_1$ where questions 2, 4, 8 and 10 refer to c_2, a_1, d_2, and c_2 respectively, we have:

$$\text{Questions } p = (1, 2, \ldots\ldots n)$$

$$W = \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \left[\begin{array}{cccccccccc} 0 & .286 & 0 & .714 & 0 & 0 & 0 & .286 & 0 & .286 \end{array}\right] \end{array}$$

$$\text{Basic Tactics } (i = 1, 2, \ldots\ldots m)$$

$$L = \text{Questions} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 4,10 & 10,4 & 2,8,10 & 4 & 8,10 & 2,8 & 2,10 & 2 & 8 & 10 \end{array} \\ \left[\begin{array}{cccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & .5 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \end{array}$$

$$\text{Basic Tactics}$$

$$\text{and } X = \begin{array}{c} \begin{array}{cccccccccc} 4,10 & 10,4 & 2,8,10 & 4 & 8,10 & 2,8 & 2,10 & 2 & 8 & 10 \end{array} \\ \left[\begin{array}{cccccccccc} 1.0 & .857 & .758 & .714 & .572 & .572 & .572 & .286 & .286 & .286 \end{array}\right] \end{array}$$

and likewise for problems built around other structures.

6) The final step analyzes the tactics for the use of redundant and/or irrelevant questions. The total amount of irrelevancy is set equal to -1.00 to be divided among the irrelevant questions of the problem. If a problem has three irrelevant questions, each one of them has a value $Ir = -0.33$.

A relevant question is redundant when it appears in a tactic with a more general question covering the same information. Instead of receiving a positive information value, all the redundant possibilities share a total weight of -0.25. If there are two possibilities as in Figure 2a for example, either c_2 or d_2 with a_1, then c_2 or d_2 with a_1 would receive a weight of $Rd = -0.125$ instead of a positive information weight as would occur if the a_1 question were not in the tactic.

The score for the observed tactic i is defined then as $S_i = x_i + Ir_i + Rd_i$, where $-1.00 \leq Ir_i \leq 0$, and $-0.25 \leq Rd_i \leq 0$ and x_i is the score of the basic tactic.

Correlations between this new scoring system and the previously used schema pulling out has been found to be, in several cases, of an order of magnitude of .90. Still the new method here described eliminates uncertainties inherent to previous procedures and has interesting theoretical and practical connotations.

General manner of evaluation

Multivariate analysis and more specifically Profile Analysis will be used throughout the study. In all cases the design can be presented as indicated in Table 3.

Table 3

General Methodological Design

		Problems			
Subjects		1	...b...	t	
	1				
	.				
	.				
	.				
Sample 1	k				
	.				
	.				
	.				
	n_1				
					⋮
	1				
	.				
	.				
	.				
Sample g	k		x_{kgb}		
	.				
	.				
	.				
	n_g				
					⋮
	1				
	.				
	.				
	.				
Sample r	k				
	.				
	.				
	.				
	n_r				

where x_{kbg} means score of subject k in sample g and problem b, $k = 1....2...n_g$, $g = 1....2...r$, and $b = 1....2...t$.

The first hypothesis to be tested corresponds to the parallelism of group profiles and can be stated as follows:

$$\begin{bmatrix} \mu_{11} - \mu_{12} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{1(t-1)} - \mu_{1t} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{gi} - \mu_{g2} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{g(t-1)} - \mu_{gt} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{r1} - \mu_{r2} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{r(t-1)} - \mu_{rt} \end{bmatrix}$$

μ_{gb} stands for the general designation for the population value corresponding to problem b in sample g.

The second hypothesis to be tested corresponds to equality of sample levels.

The third hypothesis refers to equality of problem levels.

The computational procedure used follows the method presented by Morrison (1967). Anderson (1958) refers to a similar procedure.

If parallelism is not acceptable, then the second and third hypotheses will be tested respectively by t independent univariate analyses of variance so that each variable b will be studied at the r sample levels. The hypothesis of equality of test means will be analyzed using T^2 test within each sample and for repeated measurements. Scheffe's (1953) simultaneous confidence intervals will be used to determine which one of the comparisons between samples or which one of the comparisons between tests are significantly different.

In the present study samples refer to either 1) age groups, or 2) I.Q. levels, or 3) grades, or 4) grade point averages. Problems will be studied considering two major variables: structure and language as indicated in Table 4 below.

		Structure	
		Same (Isomorphic)	Different
Language	Same	I Across years of testing for each structure and each language	II Across structure for each language and each year of testing
	Different	III Across language for each structure and each year of testing.	IV ———

Table 4

The comparisons corresponding to cell I: same structure, same language are of the type:

$31P_1 - 31P_2 - 31P_3$

$31VA_1 - 31VA_2 - 31VA_3$

$31VB_1 - 31VB_2 - 31VB_3$

$33P_1 - 33P_2 - 33P_3$

$33VA_1 - 33VA_2 - 33VA_3$

$33VB_1 - 33VB_2 - 33VB_3$

$35P_1 - 35P_2 - 35P_3$

$35VA_1 - 35VA_2 - 35VA_3$

$35VB_1 - 35VB_2 - 35VB_3$

These comparisons will indicate changes in scores as a function of first, second and third year of testing in isomorphic problems. These will be analyzed with reference to different samples. For each comparison paralellism, sample level and mean scores will be analyzed. The number of cases involved in each comparison will correspond to the number of subjects who were administered all these tests detailed in each of the comparisons tested above. The P problems deserve special consideration because in the third year they were given only to a limited group of subjects who were not able to perform problems in the VA and VB languages.

The comparisons implied in cell II: different structures, same languages, will be all performed for tests administered in the same testing session.

These comparisons are:

$31P_1 - 33P_1 - 35P_1$

$31VA_1 - 33VA_1 - 35VA_1$

$31VB_1 - 33VB_1 - 35VB_1$

$31P_2 - 33P_2 - 35P_2$

$31VA_2 - 33VA_2 - 35VA_2$

$31VB_2 - 33VB_2 - 35VB_2$

$31P_3 - 33P_3 - 35P_3$

$31VA_3 - 33VA_3 - 35VA_3$

$31VB_3 - 33VB_3 - 35VB_3$

Notice that in each comparison the language is constant but the structure varies. As in the previous case these comparisons will be performed for different samples. These comparisons will indicate the effects of structure on subject performances when language is kept constant.

The comparisons in cell III: same structure, different languages should serve to indicate the effect of language on problems built around the same structure. According to the testing design, all three languages were not regularly administered to a subject each year. Consequently the size of the samples for those comparisons are reduced in many cases. Previous evidence (Rimoldi, Aghi, Burger, 1968; Rimoldi, Vander Woude, 1969) has shown the effects of language on performance in problems with the same structure. An inspection of the overall means corresponding to these types of comparisons can be done by inspecting Table 5. Here however, though the groups are comparable, the scores do not represent repeated measures.

The comparisons in cell IV are not compatible with the hypotheses of this study.

Results*

This section will be organized in several parts. In Part I the analyses of the effects of chronological age on problem performance will be examined. In Part II the effects of different I.Q. levels will be discussed. In Part III each chronological age will be discussed separately and in Part IV problem performance will be examined in relation to grade point averages for each age group separately.

In these analyses an attempt will be made to differentiate the effect of three major variables: language of the problems (P, VA and VB), structure of the problems (31-33-35), and administration (1st-2nd-and 3rd). The complete data on which these analyses are based are presented in Tables 5 to 20.*

In Table 6 the overall mean and standard deviation for all the subjects of specified age groups are given for the three successive administrations of the picture problems. Notice that all problems with the same structure are isomorphic, except for those cases listed in the chapter on Method. In each block of Table 6 the number of subjects that were given each problem in each administration is also listed.

Tables 7 and 8 follow the same pattern as Table 6, but refer respectively to problems in the VA and VB languages. The comments made with regard to Table 6 also hold here.

Table 5 presents the means, standard deviations and number of subjects that were administered each problem in each testing session regardless of age. Tables 9 and 10 give respectively the means and standard deviations as well as the total number of subjects included in the multivariate analyses that were given problems in the P language. These are presented separately for subjects with high (H), medium (M) and low (L) I.Q. levels and for the total (T). In some cases there were not enough subjects within a specified I.Q. level to perform the analyses. Still the results for the total sample are included in the table (for instance, 8 year old sample, 2nd administration).

Tables 11 and 12 and Tables 13 and 14 are similar to Tables 9 and 10 but refer respectively to VA and VB problems.

Tables 15, 16, 17, 18, 19, and 20 are similar to Tables 9 to 14, the difference being that subjects have been classified in terms of their grade point average.

In these tables all the basic statistics pertaining to this research are included. It should be noticed that the statistical tests performed in this study require the covariance matrix corresponding to the pertinent variables. The total number of subjects used for this purpose will then depend on the number of subjects that at each age or ability level, or given grade point average (G.P.A.) were tested in "all" the problems of a specified type. This explains the discrepancies in the number of subjects as they appear in the different tables here given.

---

*Tables referred to here may be found starting on page 63.

In Tables 21 to 34 the results obtained by testing the three hypotheses indicated in the section on Methods are given. Also the significance of the simultaneous confidence intervals for different types of contrasts are presented. In Table 21 the profiles corresponding to different ages are analyzed. In Table 22 the profiles corresponding to the total sample subdivided into high, medium and low I.Q. levels are examined. In Tables 23 to 28 the profiles for the high, medium and low I.Q. groups are studied for each age group separately. In Tables 29 to 34 the same analysis is made but the basis for the comparisons is G.P.A. levels.

Each one of the Tables 21 to 34 has been arranged so that the upper half refers to cell II in Table 4. That is, the basic data is provided by the subject's scores as obtained in the same administration in problems expressed with the same language but different structures. The language of the problems used is given in the first column and the administration session in the second column of these tables.

The results of testing the three major hypotheses are indicated as follows:

1) In column 6, parallelism of the profiles. This indicates whether or not the average scores for different subsamples in problems with different structures give parallel curves.

2) In columns 3 to 5 inclusive the results indicate whether or not the profiles for each subsample are at significantly different levels.

3) In columns 7 to 9 inclusive the significance of the overall difference across structures is given together with the specific contrasts.

Given that the profiles are parallel, the results of testing the second hypothesis indicate whether the classification into the defined subsamples does make for significant differences in scores. The results of testing the third hypothesis indicate whether the different structures give results that are statistically different.

The lower half of the tables refers to problems with the same language and the same structure in different administrations. Thus columns 3 to 5 inclusive give results that indicate whether or not the profiles corresponding to the subsamples used differ significantly. Columns 7 to 9 inclusive indicate whether or not administration produces a significant change in scores.

Notice that the cells corresponding to columns 3 to 5 inclusive and 7 to 9 inclusive in Tables 21 to 34, two sets of values are given. Those in the upper half refer to the overall evaluation of the contrast. Those in the lower half refer to specific comparisons as indicated at the top of the corresponding columns.

With reference to the contrasts studied it should be remembered that the Scheffé's confidence intervals give very conservative measures.

In Appendix B a graphical representation of the content of Tables 21 and 22 is given. This supplementary material is provided to help the reader visualize some of the results described below. For the tables presented here it is possible to construct similar graphs.

Part I

In this part the results obtained are discussed when studying the effects of age levels on problem performance. The samples used consist of those subjects examined for the first time at 7, 8, 9, 10, 11, and 12 years of age. These samples will be identified by a 7 year old sample, an 8 year old sample, a 9 year old sample, etc. These subjects will be one year older in the second administration and two years older in the third administration, so that the 9 year old sample in the third administration is of the same age as the 11 year old sample in the first administration. The information corresponding to this part is presented in Table 21.

The main points to be discussed will be presented as a series of questions 1 to 4. These questions refer to how age levels relate to performance considering language, structure and administration of the problems (1st, 2nd or 3rd administration). In all cases we have made comparisons based on all the possible pairs of ages, as well as contrasts based on pooling together different age groups.

For all the results here reported the profiles corresponding to specified age levels are statistically parallel.* This means that the relative differences in performance are the same for different age levels.

Question 1

How do scores in problems with different structures and same language vary according to chronological age and administration (1st, 2nd or 3rd year of testing)? These changes will be analyzed for each language separately.

With regard to problems 31P, 33P, and 35P in the first administration the overall significance for different age levels is at the .01 level, that is problems in language P do differentiate between age groups, regardless of the structure used. More specifically these differences separate those subjects of 7, 8 and 9 years of age from those of 10, 11 and 12 years at the .01 level. While there is no difference between any pairs of the ages 10, 11 and 12, or any pairs of the ages 7, 8 and 9, the comparisons (7-10) (8-10) and (9-10) are significant as indicated in Table 21. Since the older group gives results that are significantly greater than those of the younger group, it may be suggested that between 9 and 10 years of age occurs a definite change in level of performance for the P problems.

For the same problems among the second administration, only subjects in the 7 and 8 years of age samples were tested (at this administration the original 7 year old group is 8 years of age and the original 8 year old group is 9 years of age). The reduction in sample size with regard to the first year of testing is due to the order followed in the test administration. As can be seen in Table 21 the two groups are significantly different at the .05 level with the 7 year old sample giving higher scores than the 8 year old sample. It should be considered that at this second administration the only 8 year olds that were given problems in the P language were those that performed badly in the VA and VB problems. Thus the fact that their performance is significantly lower than for the younger group is not surprising, and on the contrary it was expected.

No comparisons are possible at the third year of testing since only the 7 year old sample (now of 9 years of age) was administered problems in P language.

---

* Parallelism was tested at the .05 and .01 level using D. L. Heck charts as presented by Morrison, 1967.

With reference to the VA problems during the first administration, they were given to the 8 to 12 year old samples. The overall results indicate that the profiles corresponding to the different age groups have a significant difference in levels (.05). More specifically the significant differences at .10 level, correspond to the comparisons (10-12) and (10-11 and 12). That is, the difference in levels seems to occur between 10 and 11 years of age. Notice that the 8 year old group (represented by only five subjects) gives higher results than any of the other age samples. This may be explained considering the selection procedure used in the test administration (see Method), that is these 8 year olds performed better than most of their peers in the P problems so that they were given the VA problems.

During the second administration of the VA problems, subjects from the 7, 8, 9 and 10 year old samples (now of 8, 9, 10 and 11 years of age) were used. The differentiation between age levels is weak and separates the 10 year old sample from the 9 and 8 year old samples.

The third administration included 7, 8, 9 and 10 year old sample, now of 9, 10, 11 and 12 years of age respectively. The differentiation between age levels is at the .01 level, and occurs in the comparisons (7-9), (7-10), (8-9), (8-10) and (7, 8-9, 10). This indicates that a difference in performance occurs between 10 and 11 years of age as was also the case in the previous administration. In summary all the comparisons performed seem to indicate that at about 10 to 11 years of age there is definite improvement in performance in the VA problems so that two relatively homogeneous subsamples could be defined below and above this age level.

Only subjects of 9, 10, 11 and 12 years of age were tested during the first administration in VB problems. The overall test of significance shows a differentiation between ages at the .05 level. It should be noticed that the 9 year old sample gave higher scores than the 10 year olds. The difference is significant at .05. Again, because of the order in which the problems were administered, the 9 year olds for this analysis represent only approximately 50% of the 9 year olds who were capable of handling the verbal problems. The comparison (10-11, 12) is significant at the .05 level.

During the second administration the VB problems differentiate at the .05 level between age groups. This overall difference is due to the 8 year old sample and 10 year old sample (now of 9 and 11 years of age respectively).

At the third administration the VB problems differentiate age levels at the .01 level. The samples studied are 7, 8, 9, and 10 which now have 9, 10, 11 and 12 years of age respectively. The picture is similar to the one encountered with the VA problems. Inspection of Table 21 shows that tne comparisons (8-9), (8-9), and (7-10), are significant at the .05 or .01 level, while the comparison (7-9) is only significant at the .10 level. Pooling together the 7 and 8 year old samples and the 9 and 10 year old samples, a contrast that is significant at the .01 level is obtained. That is no difference between 7 and 8 year old samples or between 9 and 10 year old samples but a significant change between the 8 and 9 year old samples, now respectively of 10 and 11 years of age. Again the suggestion is that between 10 and 11 years of age there does occur a significant change in performance.

Thus, while problems in language P suggest a change between 9 and 10 years of chronological age, for problems in languages VA and VB this change occurs somewhat later, that is between 10 and 11 years of age. It would be of interest to know how at later ages problems in P, VA and VB languages behave. According to our previous experience, (Rimoldi, H., Aghi, M., Burger, G., 1968) it may be hypothesized that the increase will be greater for the VA problems (and perhaps also for the VB problems) than for the P problems.

It also looks, judging from our results, that for P problems the differentiation is sharper at younger than at later ages, while the opposite seems to occur with the VA and VB languages. Therefore presentation of structures in VA and VB languages will differentiate better age levels in subjects above 10 or 11 years of age, while presentation of the same structures in P language will be more effective to differentiate earlier age levels.

Question 2

How do scores in the problems differentiate between structures 31, 33 and 35?

The results given in the upper part of columns 31-33, 31-35 and 33-35 of Table 21 provide the pertinent answer. The consistency of the results is worth noticing. Regardless of the language used, the different structures give overall statistically significant results. In all cases but one (problems in P language, first administration) structure 33 is significantly different from structure 35. The difference between structures 31 and 33 is significant only for problems in P language, first administration. With reference to structures 31 and 35, in all cases the difference is significant at the .01 level, except for problems $31VA_1$ and $35VA_1$ (significant only at the .10 level) and problems $31P_1$ and $35P_1$ (significant at the .05 level). These results were obtained by pooling together the age groups indicated in the columns at the right of Table 21.

Inspection of the corresponding means (Table 6, 7 and 8) indicate that the scores in structure 35 are lower than those corresponding to structure 33. Similarly structure 31 gives always higher scores than structure 35 regardless of language used or administration (first, second or third). Also there seems to be no difference between structures 31 and 33, except for problems in P language, first administration, where structure 31 gives higher scores than structure 33.

Considering that the scoring system used does not depend on sample norms, but relates to the logical structure of the problem, these findings suggest that more than a purely ordinal evaluation of problem difficulty has been reached, provided the rationale used for developing the scoring system is accepted.

In summary the comparisons over several age groups and language differentiate logical structures. In this study the indication is that structures 31 and 33 are "easier" than 35, with structure 31 and 33 of about the same difficulty level. The already noted exception refers to problems $31P_1$ and $33P_1$ This may be due to the fact that this is the first administration. Also, the sample used has a large number of 7 year old children who might be able to understand significantly better structure 31 than structure 33 when both are presented in a perceptual context. At later ages or in successive administrations and with other languages, this difference tends to vanish as it may be indicated by our results.

This suggests that for evaluating cognitive ability as well as for determining when a certain concept should be presented within an education program, it might be worth considering: a) the logical structure of the concepts and b) most effective manner of presentation. Though logical structures and languages are within limits operationally independent, psychologically considered they probably interact, so that the language used to express a certain structure is not a matter of complete indifference.

Question 3

What changes in scores do occur in problems with the same language and the same structure for different age levels?

The corresponding findings are given in the lower part of Table 21.

Since P problems were administered only to the 7 year olds for three successive years, no comparisons with other age groups are possible. Regarding the VA problems, graphs were made for problems of same structure, and same language in the three successive administrations, separating them into age levels. The curves corresponding to the different age samples are parallel and their differences in levels are not statistically significant. In all cases the subjects of the lower age group give higher scores than those of successively greater ages. This result is probably an artifact due to the order in which the problems were administered.

For the VB problems a similar result is obtained. Only two age groups were examined (8 year old and 9 year old samples) and the differences in level are significant at the .10 level.

Question 4

What changes in scores occur in problems with the same structure and the same language over years of testing (first, second and third administration)?

Here it is clear that in all cases administration does produce a significant change, whether the problems have structures 31, 33 or 35 or whether they are presented in the VA or VB languages. The findings show that a significant difference occurs between first and third and between second and third administration. The scores increase from first and second to third administration.

Part II

In this part of the report, we hall analyze the results obtained when subjects are classified into high, medium and low in terms of their I.Q. levels. This classification involved all the subjects of a given age; that is, the top one third was the high group, the middle third was the medium group and the lower third the low group. The results referred to the total sample ignoring age levels so that the statements made do not differentiate between different ages. We shall compare the results for different structures and languages as well as for different administrations. The results are given in Table 22. Though I.Q. levels vary slightly between samples, they are highly similar as can be seen by inspecting this table.

Notice that within any of the comparisons made, the corresponding profiles are always parallel, so that the curves for the high, medium and low groups show changes that are statistically similar.

Question 1

How do problems with different structures but expressed in the same language differentiate in terms of I.Q. levels and administration (first, second and third administration)?

For the problems expressed in P language, the difference between I.Q. levels is significant at the .05 level during the first administration and only at the .10 level in the third administration. The results for the second administration are not statistically significant.

In most of these comparisons the high group shows scores that are higher than the medium group which in turn shows higher values than the low group. The only comparison that is significant for the P problems refers to the contrast between H and L group. That is, the P problems differentiate clearly between H and L groups during the first administration, but weakly during the third administration (.10 level). Since the sample of subjects that received the P problems is heavily loaded with subjects that below to the younger ages, this seems to indicate that the P problems differentiate mostly I.Q. levels in subjects at the lower ages. This is in general agreement with the previously discussed findings.

With reference to the VA and VB problems the results are highly consistent. In the first place the overall level of significance is at the .01 level for all the administrations in all the problems. Secondly the high and low groups are also always significantly different at the .01 level. The difference between the high and medium group is always significant at that same level, except for problems of the VA series in the two administrations (.05 level) and for problems of the VB series in the third administration (.10 level). The contrast between the medium and low groups is significant at the .01 or .05 levels except for the second administration.

In summary, the P problems differentiate between extreme I.Q. levels mostly during the first administration (in which the 7 and 8 year olds represent 65% of the total sample). In the second and third administrations, this difference is weak or non-existent.

On the other hand the VA and VB problems do differentiate between any of the possible combinations of I.Q. levels in all administrations, except for the VA and VB problems in the second administration, which nevertheless approach significance.

Notice that the highest level of significance for the three I.Q. comparisons occur in problems of the VA and VB types in the first administration. In this administration the age groups 10, 11 and 12 year olds represent approximately 85% of the total sample (see Table 21). It seems reasonable to infer that the VA and VB problems do differentiate most effectively between I.Q. levels at the high age groups while the P problems are most effective in differentiating I.Q. levels at the younger ages.

Question 2

How do scores in the problems differentiate between structures?

As shown in Table 22 the results are highly consistent throughout different languages and administrations. The overall significance between structures is always at the .01 level, except for problems in the VA language during the first administration (.05 level). The general pattern is quite similar to the one discussed when studying age levels (Table 21). This in itself is an interesting finding since in spite of the fact that the subjects involved are the same (when studying age levels and when studying I.Q. levels) the particular pattern of significant differences did not have necessarily to agree. The fact that this has happened is to be noticed. It seems to indicate that regardless of the two classification systems used, structures show the same differentiating power.*

Question 3

What changes do occur in problems in the same language and with the same structure as a result of different I.Q. levels?

This implies studying the performance in three problems with the same language and same structure through three successive administrations. In the first place neither problems of the 31P or 33P series differentiate among I.Q. levels. Problems of the 35P type differentiate only at the .10 level between the high and low groups. The pattern for the VA and VB problems are highly similar. Whether they correspond to structures 31, 33 or 35, all of them give overall significant differences between I.Q. levels (.01) and in all cases the comparison H-L is also significant at that level. Only problems of the 31VA type differentiate between medium and low I.Q. levels (.01). The contrast H-M is significant at the .01 level for all problems except for those of series 31VA and 33VB.

---

* Note that the number of subjects in Table 21 are very slightly lower than those in Table 22. When the classification of the subjects was made with regard to age, those age groups with N $\leq$ 2 were not included.

These findings summarize the results of the comparisons based on the three administrations of problems with the same structure and same language and indicate that within the same administration, performance in the problems relate to I.Q. levels so that subjects with higher I.Q.'s will in general perform better than those with a low I.Q. level. It should be remembered that when the same type of comparisons were performed with regard to age levels, the results were essentially negative.

It is worth noticing that with respect to Question 1, the significance between levels was due to the H-L, H-M, and L-M comparisons throughout different structures. In the case that we are now discussing the contrasts are analyzed through administrations and the significant M-L comparison disappears (except for problems of the 31VA type). Again it seems that for the purposes of differentiation among I.Q. levels it may be more sensitive to compare subjects performances in terms of several structures than to administer several isomorphic problems in succession.

Question 4

What changes in scores occur in problems with the same structure and language as a result of administration, when these changes are referred to I.Q. levels?

For the VA and VB problems the results are almost exactly the same as those reported in Table 21. The P problems follow the same pattern that the VA and VB problems follow, except for those of the 31P series. Here the overall difference is at the .05 level and due primarily to the contrast between the second and third administration.

In all cases performance in the third administration is higher than in the first or second administration. The second administration may or may not give higher scores than the first administration.

Part III

In this part we shall analyze the results corresponding to each age group separately with regard to I.Q. levels. The influence of structures and years of testing will be discussed.

By age group we mean the age of the subject the first time that he was tested. Thus the chronological age at a particular administration corresponds to the age group plus one or two years respectively for the second and third years of testing. For each age group a table will be presented summarizing the results obtained.

7 year old sample (Table 23)

The results are based on those subjects who were administered the three problems used for comparison, that is either isomorphic problems through the three years of testing or the three different structures presented in the same language. Because of these conditions as well as because of the order in which the problems were administered the total number of subjects in each comparison is not constant.

With reference to problems based on different structures but expressed in the same language, we find that in all cases (except for the problems 31P$_1$, 33P$_1$ and 35P$_1$) subjects' performance differentiates weakly between I.Q. levels and that this is only due to the contrast between the high and low I.Q. groups. The significance of this contrast is higher for the VA problems in the third administration, that is when the average age of the subject tested was 9 years of age.

Notice that for all the comparisons the profiles corresponding to the three I.Q. levels are parallel. This implies that the changes observed among the subjects belonging to a specified I.Q. level are not statistically significant from the changes corresponding to subjects at a different I.Q. level.

With regard to structures the same pattern of significant values that is shown is Table 22 has been found. That is, the overall significance reaches the .01 level in all the problems administered to the 7 year old sample except for the P problems, third administration. More specifically, the significant differences refer to the comparisons between structures 31 and 33 and 33 and 35. The difference (31-33) is significant at the .01 level only in problems presented in P language, first administration. The lack of significance in P problems, third administration, indicates that at this age level (9 years of age) subjects that have been tested twice before, perform equally well on all the structures involved in the comparison. It is interesting to notice that this does not occur for problems of the VA and VB type during the same third administration. The possibility that this lack of differentiation with the P problems may result from a leveling off in performance (plus the influence of possible training effects) may not be dismissed. That is, the subjects at this age do understand the three structures involved in the three problems and can operate equally well in any one of them if presented in P language.

Comparing now problems with the same structure and the same language at different administrations, it is found that across administrations there are no significant differences due to I.Q. levels in P problems with structure 31 and 33. However, with structure 35 the results indicate that the subjects with high I.Q. level perform better (.10 level) than those with a low I.Q. level.

Comparing now level of performance during the three successive administrations, we found that for all the structures the third administration gives results significantly better than the first or second testing sessions except for problems in the 31 series. That is, scores in structure 31 expressed in P language change little with subsequent administrations.

Notice that for the 7 year old sample the cells corresponding to first and second administration in VA and VB languages are empty. This is due to the fact that very few of the 7 year olds were administered these problems either the first or the second year of testing. For the same reason the comparisons of problems in VA and VB languages with the same structure but different administrations could not be performed.

In summary, the general comments made in the previous sections of this study seem to apply to the 7 year old sample.

8 year old sample (Table 24)

Only very few subjects of this sample were administered the P problems in the second and third administration, that is when their chronological ages were respectively 9 and 10 years of age. At the same time they received VA and VB problems during the second and third administration.

Comparing performances in problems $31P_1$, $33P_1$, and $35P_1$, for the three different I.Q. levels, we observe that the overall significant difference occurs at the .10 level and is due only to the comparison of the high and low I.Q. groups. For the VA and VB problems the comparison of I.Q. levels gives an overall significance at the .01 level in all cases, as well as for the specific high-low contrast.

The comparisons medium-low are only significant at the .10 level and occur in both VA and VB problems during the second administration. The high-medium contrast is significant at the .01 level in problems of series VA_3, otherwise the contrast is weak. That is, the pattern is quite similar to the one reported for the previous analyses.

For all the problems administered to the 8 year old sample the overall comparison through structures is always significant. The differences between structures 31-35 are in all cases significant at the .01 level, except for problems in language P, first administration, where the overall difference is significant at the .10 level and due to the contrast between structures 31 and 33. Again as in previous cases, structure 33 is significantly different from structure 35 in the VA problems in the second and third administrations.

As can be observed in the corresponding table for the 8 year sample, there were no comparisons made in any problem during three successive years. This was a result of the design used in the administration of the problems. Again the general pattern agrees with the previous observations.

9 year old sample (Table 25)

As can be seen in the corresponding table the P problems during the first administration do not differentiate in terms of I.Q. levels. The VA problems differentiate I.Q. levels in the first (.10) level) and third (.05 level) administrations. The VB problems differentiate I.Q. levels only in the second administration (.05 level). Notice that at this age level the P problems were only administered if the subjects performed poorly in the problems administered previously. This contributes to the explanation of the lack of differention of I.Q. levels found with the P problems. For the VA problems during the first administration, the I.Q. level differentiation is weak, and for the VB problems it is not significant. Nevertheless, for the 8 year old sample, second administration (9 years of age) the results were highly significant. It may be postulated that the VB problems are too difficult at this age level while with the VA language the subjects may have some greater facility. Still this does not explain why the VA_2 problems (that correspond to 10 years of chronological age) and the VB_3 (that correspond to 11 years of chronological age) do not differentiate between I.Q. levels. The inspection of the means involved in these contrasts does not help much, and we can not now offer an explanation.

With regard to structures, the same pattern that was previously found becomes again quite apparent. The comparison is always significant for the contrasts 31-35 and 33-35. In the case of the P problems, it is only the comparison 31-33 that reaches a .05 level of significance.

For problems with the same structure and language it is found that those of the series 31VA do not separate between I.Q. levels, while those of series 33VA and 35VA do. The profile for problems 33VA was not parallel and thus three univariate analyses of variance were performed showing that it is only at the third administration (11 years of chronological age) that the structures are significantly different. This can be readily seen by inspecting the corresponding graphs.

While problems of the series 33VB and 35VB do not differentiate between I.Q. levels, problems of the series 31VB do. Again the profiles are not parallel and the analyses of variance show that during the first administration, I.Q. levels are significantly different at the .01 level, while at the third administration the level reached is only .10.

With regard to years of testing, notice that the pattern previously found is repeated here. This is, scores increase from first to second to third year of testing. In the comparisons in which non-parallel profiles are involved, that is VA33 and VB31, Hottellings' T^2 test shows that the changes in scores with administration are significant at the .01 level for the medium and high groups. For problems of the series VB31 the low I.Q. group shows significant changes between administrations.

With regard to this 9 year old sample the results share many features in common with those previously found. However, the picture is less sharp than for younger or older groups. The subjects here studied are those that were originally of 9 years of age. At the second and third administration their ages will be respectively 10 and 11 years of age. As previously hinted, it is at this age level that some basic changes seem to occur with regard to the effective use of the languages involved in this research and with the understanding of the structures. Therefore the findings here reported do show the effects that such a transition period is most likely to produce.

10 year old sample (Table 26)

Only a small sample of 10 year olds had to be administered problems 31P, 33P and 35P. Inspection of the corresponding graphs shows a high level of performance for the profiles corresponding to the three I.Q. levels which are not significantly different. Nevertheless the VA and VB problems show in all cases (first, second, or third administration) differences in the profiles corresponding to different I.Q. levels. These are significant at the .01 level. This is mostly due to the contrast high-low I.Q. The second most predominant contrast refers to the comparison H-M, especially noticeable in the second and third administration when the subjects were respectively of 11 and 12 years of age. Notice that the comparison medium-low I.Q. is significant in both VA and VB problems during the first administration. The similarity of these patterns of differences with the overall pattern indicated in Table 22, is intriguing in the sense that while in the majority of the cases the significance of the differences between I.Q. levels depends mostly on the H and L I.Q. groups, it is only at the older age levels that the differences between high and medium I.Q. groups and medium and low I.Q. groups begin to emerge. This of course relates to the differentiating power of the structures and languages used and their relation to age.

At this age level the P problems do not differentiate among structures, the implication being that for all the problems used the performance is at a high level regardless of the structure used. Nevertheless for the VA and VB problems the results indicate not significant differences with regard to structure during the first administration. During the second and third administration the overall test of significance shows that structures are differentiated at the second and third administration. As always the predominant contrasts refer to structures 31-35 and 33-35 repeating the overall findings given in Table 22. For this 10 year old sample second and third administration correspond respectively to 11 and 12 years of age.

Comparing problems with the same structure and language in the three administrations, it is found that in all cases the comparisons are significant (.01) and that the specific contrast high-low I.Q. is also significant at the same level. For the VB problems the differentiation is also significant between the high and medium I.Q. groups in all the structures.

For the VA problems, structures 33 and 35 behave similarly.

With regard to administration notice that except for problems VB 31, all the others provide significant results indicating that there is always a significant difference in the scores obtained in the first and third administration as well as in the second and third ones.

The picture given by the 10 year olds is definitely clearer than the one shown by the 9 year old group. It appears that at this level (which includes subjects of 10, 11 and 12 years of age) differentiation of I.Q. levels, structures, administration and languages becomes well delineated.

11 year old sample (Table 27)

Problems VA and VB were administered only once according to the design (Table 1). Thus comparisons through differnt testing sessions will not be made.

The results for problems of the VA type and VB type indicate that they differentiate successfully between I.Q. levels (.01) and that the H-L contrast is significant at that same level.

At this age level, the performance in the different structures gives significant results (.05 level) due mostly to the significant contrast between structures 31 and 35. Notice that the scores for these problems are high.

12 year old sample (Table 28)

The general comments made with regard to the 11 year old sample also hold here. The results repeat the same general picture already discussed. The VA problems differentiate between I.Q. levels (.01). The significant contrasts refer to the comparisons between high-medium, and high-low I.Q. levels. The VB problems differentiate similarly, including the comparison medium-low that is significant at the .10 level.

With regard to structures, problems 33 and 35 are clearly differentiated in both VA and VB problems. The contrast 31-35 is significant (.05) only in problems in the VB language.

structures better than grade point average, at the second and third year of testing (9 and 10 years of age respectively). Otherwise the patterns are highly similar.

9 year old sample (Table 31)

For the 9 year old sample the contrasts between structures are the same in terms of G.P.A. and I.Q. levels. Similarly with regard to contrasts involving administration in problems with the same language and same structure (isomorphic problems). Notice that the lack of parallelism found for VA33 and VB31 disappears with using G.P.A. as a criterion for classification. Columns 3 to 5 inclusive in Table 31 show that for P problems there is no difference in the samples based on I.Q. levels or in G.P.A. levels. For the VA problems the classification in terms of G.P.A. scores differentiates better than the I.Q. level classification when we consider the contrasts that involve problems with the same language and different structures. And a similar result is observed for the VB language.

Comparison of isomorphic problems through administration does not give a clear picture. At this age level, the pattern that had prevailed in previous years seems to be shifting in the sense that VA problems become better discriminators in profiles based on G.P.A. scores than in profiles based on I.Q. levels. It is appropriate to remember that the 9 year old sample includes subjects of 9, 10 and 11 years of age for the first, second and third administration. It should also be remembered that it was at the 9 year old sample that the pattern of contrasts found for I.Q. levels was not clear. In summary it seems that this age sample corresponds to a transition period in which the increasing influence of VA language seems to be the predominant factor. Apparently the P language has here less differentiating power.

10 year old sample (Table 32)

For the 10 year old sample, classification of subjects in terms of G.P.A. scores gives substantially the same results as classification in terms of I.Q. levels. The most noticeable difference is that in contrasts involving the same language and different structures as well as in contrasts involving isomorphic problems, there is a predominence of significant differences between high and medium and medium and low levels when classifying subjects in terms of grade point averages. For the I.Q. classification the differences were mostly in terms of high-low and high-medium levels. This may suggest that grades in school tend to differentiate more sharply among the lower achievers while I.Q. will separate more clearly subjects in the high and medium levels.

11 year old sample (Table 33)

For the 11 year old sample, the picture given by the profiles based on G.P.A. levels is almost exactly the same as the one given by the I.Q. profiles.

12 year old sample (Table 34)

As in the case of the 11 year old sample, no major differences are found between classifications based on I.Q. and G.P.A. levels.

Table 5. Total Sample - Ms, σs, and Ns*

Lang.		1st Year				2nd Year				3rd. Year				Average			
		31	33	35	Ave.	31	33	35	Ave.	31	33	35	Ave.	31	33	35	Ave.
P	M	.52	.42	.45	.46	.49	.52	.27	.43	.64	.66	.57	.62	.52	.47	.42	.47
	σ	.34	.38	.44		.29	.40	.40		.22	.23	.28					
	Ns	305	298	290	893	110	110	110	330	40	40	40	120	455	448	440	1343
VA	M	.64	.66	.64	.65	.51	.61	.48	.53	.70	.70	.57	.66	.62	.66	.58	.62
	σ	.32	.30	.31		.34	.28	.40		.29	.30	.35					
	Ns	492	468	447	1407	298	291	287	876	354	350	338	1042	1144	1109	1072	3325
VB	M	.51	.50	.42	.48	.45	.44	.35	.41	.54	.52	.41	.49	.50	.49	.40	.46
	σ	.32	.33	.35		.33	.33	.33		.32	.36	.37					
	Ns	430	422	415	1267	281	280	281	842	319	319	317	955	1030	1021	1013	3064
Ave.	M	.56	.54	.51	.54	.48	.53	.39	.47	.63	.62	.50	.58	.56	.56	.48	.53
	σ																
	Ns	1227	1188	1152	3567	689	681	678	2048	713	709	695	2117	2629	2578	2525	7732

*"N" for any category designated "Ave." represents number of problems, while "N" elsewhere represents number of different subjects.

Table 6. Picture Problem Means, Standard Deviations, and Number for the Total Sample

Age at Different Administrations

Age at First Administration		7			8			9			10			11			12		
		31	33	35	31	33	35	31	33	35	31	33	35	31	33	35	31	33	35
7	M	.49	.36	.36	.50	.57	.29	.65	.66	.58									
	σ	.37	.37	.45	.28	.38	.40	.22	.24	.28									
	n	105	102	97	91	91	91	39	39	39									
8	M				.48	.40	.42	.44	.32	.20	.44	.60	.25						
	σ				.32	.39	.45	.34	.42	.42	.00	.00	.00						
	n				99	96	94	18	18	18	1	1	1						
9	M							.50	.36	.46	.58	-.36	.16						
	σ							.29	.38	.42	.00	.00	.00						
	n							46	46	45	1	1	1						
10	M										.63	.64	.67						
	σ										.29	.36	.33						
	n										35	34	34						
11	M													.58	.69	.52			
	σ													.35	.25	.50			
	n													11	11	11			
12	M																.79	.59	.69
	σ																.26	.22	.22
	n																9	9	9

Table 7. Verbal A Problem Means, Standard Deviations, and Number for the Total Sample

Age at Different Administrations

Age at First Administration		7			8			9			10			11			12		
		31	33	35	31	33	35	31	33	35	31	33	35	31	33	35	31	33	35
7	M	.23	.94	1.00	.54	.62	.32	.62	.64	.43									
	σ	.53	.00	.00	.39	.19	.39	.30	.31	.38									
	n	2	1	1	12	11	8	86	82	72									
8	M				.55	.77	.71	.52	.58	.40	.65	.64	.54						
	σ				.33	.24	.19	.37	.30	.40	.35	.32	.34						
	n				11	7	5	84	79	76	84	84	82						
9	M							.55	.58	.66	.59	.55	.45	.78	.75	.61			
	σ							.37	.35	.26	.32	.29	.38	.24	.27	.34			
	n							100	85	73	89	89	89	86	86	86			
10	M										.60	.60	.58	.59	.67	.55	.74	.75	.65
	σ.										.34	.32	.36	.33	.27	.40	.26	.30	.30
	n										122	121	115	110	109	111	94	94	94
11	M													.71	.69	.65	.81	.64	.79
	σ													.25	.27	.31	.22	.34	.19
	n													130	127	128	3	3	3
12	M																.69	.73	.67
	σ																.31	.22	.28
	n																127	127	125

Table 8. Verbal B Problem Means, Standard Deviations, and Number for the Total Sample

Age at Different Administrations

Age at First Administration		7			8			9			10			11			12		
		31	33	35	31	33	35	31	33	35	31	33	35	31	33	35	31	33	35
7	M	.00	.00	.00	.32	.42	.25	.50	.44	.28									
	σ	.00	.00	.00	.25	.46	.32	.28	.33	.37									
	n				7	7	7	54	53	53									
8	M				.63	.73	.71	.39	.32	.31	.48	.42	.33						
	σ				.23	.10	.41	.31	.29	.30	.35	.38	.32						
	n				3	3	2	73	73	73	81	83	82						
9	M							.54	.57	.48	.44	.45	.34	.61	.55	.46			
	σ							.30	.32	.37	.33	.33	.35	.31	.37	.37			
	n							66	62	60	88	88	88	86	85	85			
10	M										.44	.41	.37	.51	.50	.37	.55	.63	.49
	σ										.35	.38	.34	.33	.33	.33	.33	.33	.37
	n										109	109	106	110	109	110	94	94	93
11	M													.52	.50	.44	.60	.83	.61
	σ													.30	.33	.34	.57	.17	.25
	n													127	125	123	3	3	3
12	M																.53	.52	.43
	σ																.31	.29	.37
	n																125	123	124

Table 9. Picture Problem Means and Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H	.57	.42	.42	39	.58	.66	.39	33	.67	.75	.76	9												
	M	.52	.35	.40	26	.47	.51	.25	27	.60	.69	.56	12												
	L	.42	.29	.25	32	.45	.52	.21	31	.67	.59	.51	18												
	T	.51	.36	.36	97	.50	.57	.29	91	.65	.66	.58	39												
8	H					.46	.54	.54	29				2				1								
	M					.48	.41	.48	34				3				0								
	L					.53	.23	.28	32				14				0								
	T					.49	.39	.43	95	.43	.31	.17	19				1								
9	H									.51	.34	.48	10				1								
	M									.57	.48	.45	14				0								
	L									.50	.31	.43	23				0								
	T									.52	.37	.45	47				1								
10	H													.76	.70	.94	6								
	M													.59	.52	.73	9								
	L													.63	.67	.56	19								
	T													.64	.64	.68	34								
11	H																				2				
	M																				1				
	L																				8				
	T																	.58	.69	.52	11				
12	H																								
	M																								
	L																								
	T																					.79	.59	.69	9

Table 10. Picture Problem Standard Deviations for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H	.39	.36	.45	39	.26	.29	.41	33	.21	.25	.22	9												
	M	.34	.37	.41	26	.25	.36	.36	27	.25	.19	.28	12												
	L	.34	.39	.45	32	.31	.44	.40	31	.18	.24	.26	18												
	T	.37	.38	.44	97	.28	.37	.40	91	.21	.23	.28	39												
8	H					.37	.39	.41	29	.29	.22	.40	2				1								
	M					.35	.39	.38	34	.26	.30	.38	3				0								
	L					.25	.35	.47	32	.38	.32	.35	14				0								
	T					.33	.40	.44	95	.32	.40	.42	19												
9	H									.36	.35	.29	10				1								
	M									.29	.41	.49	14				0								
	L									.20	.34	.42	23				0								
	T									.27	.37	.42	47												
10	H													.21	.46	.10	6								
	M													.39	.43	.31	9								
	L													.24	.26	.33	19								
	T													.29	.36	.33	34								
11	H																				2				
	M																				1				
	L																				8				
	T																	.33	.24	.48	11				
12	H																								
	M																								
	L																								
	T																					.25	.21	.20	9

Table 11. Verbal A Problem Means for Multivariate Analyses
Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				0				0	.68	.74	.58	34												
	M				0				2	.70	.66	.37	19												
	L				1				6	.58	.61	.20	20												
	T				1	.59	.65	.32	8	.66	.68	.42	73												
8	H								0	.63	.70	.49	26	.79	.76	.69	27								
	M								4	.61	.57	.43	30	.69	.73	.55	30								
	L								1	.45	.49	.15	20	.47	.47	.35	26								
	T					.72	.90	.71	5	.58	.60	.40	76	.65	.66	.53	83								
9	H									.71	.64	.74	30	.61	.57	.59	29	.84	.87	.68	30				
	M									.70	.63	.68	25	.57	.54	.45	30	.74	.67	.61	28				
	L									.53	.59	.49	19	.59	.55	.33	30	.74	.69	.52	28				
	T									.66	.62	.66	74	.59	.55	.45	89	.78	.75	.61	86				
10	H													.71	.72	.75	42	.62	.73	.72	41	.82	.87	.83	37
	M													.71	.59	.58	29	.69	.70	.55	25	.74	.61	.58	22
	L													.49	.54	.43	43	.51	.56	.40	45	.67	.70	.51	36
	T													.63	.62	.58	114	.59	.65	.56	111	.74	.75	.65	95
11	H																	.79	.76	.72	48				
	M																	.72	.74	.69	36				
	L																	.64	.56	.53	43				
	T																	.72	.69	.65	127				
12	H																					.76	.81	.76	49
	M																					.67	.66	.62	44
	L																					.60	.72	.60	32
	T																					.69	.73	.67	125

Table 12. Verbal A Problem Standard Deviations for Multivariate Analyses
Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at Different Administrations

Age at First Administration

		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				0				0	.25	.26	.31	34												
	M				0				2	.22	.30	.41	19												
	L				1				6	.30	.21	.37	20												
	T					.26	.20	.37	8	.26	.27	.39	73												
8	H								0	.29	.22	.40	26	.22	.23	.28	27								
	M								4	.26	.30	.38	30	.31	.20	.27	30								
	L								1	.38	.32	.35	20	.42	.37	.41	26								
	T					.12	.08	.17	5	.32	.29	.40	76	.35	.31	.35	83								
9	H									.27	.25	.26	30	.31	.26	.39	29	.18	.18	.35	30				
	M									.25	.41	.25	25	.38	.34	.37	30	.29	.33	.37	28				
	L									.28	.28	.19	19	.25	.26	.34	30	.23	.25	.28	28				
	T									.28	.32	.26	74	.32	.29	.38	89	.24	.27	.34	86				
10	H													.27	.31	.29	42	.29	.29	.32	41	.24	.22	.20	37
	M													.27	.27	.39	29	.24	.22	.39	25	.28	.41	.29	22
	L													.31	.30	.33	43	.39	.29	.40	45	.23	.22	.30	36
	T													.30	.31	.36	114	.33	.29	.38	111	.26	.30	.30	95
11	H																	.21	.18	.25	48				
	M																	.21	.21	.25	36				
	L																	.25	.35	.38	43				
	T																	.23	.27	.31	127				
12	H																					.25	.17	.26	49
	M																					.36	.26	.27	44
	L																					.30	.20	.26	32
	T																					.31	.22	.27	125

Table 13. Verbal B Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H								0	.57	.51	.35	28												
	M								2	.51	.40	.26	13												
	L								5	.38	.30	.16	12												
	T					.32	.42	.25	7	.51	.44	.28	53												
8	H								0	.53	.44	.43	26	.58	.55	.38	27								
	M								1	.39	.26	.35	29	.52	.41	.40	31								
	L								1	.22	.25	.09	19	.32	.29	.21	25								
	T									.40	.32	.31	74	.48	.42	.34	83								
9	H									.61	.57	.54	27	.52	.54	.44	29	.64	.66	.52	30				
	M									.63	.62	.44	20	.42	.49	.35	30	.57	.55	.47	28				
	L									.36	.52	.39	14	.38	.32	.23	29	.60	.44	.38	26				
	T									.56	.58	.47	61	.44	.45	.34	88	.61	.56	.46	84				
10	H													.56	.54	.48	41	.63	.66	.55	41	.70	.75	.66	37
	M													.39	.35	.37	26	.43	.42	.24	25	.54	.61	.46	21
	L													.37	.33	.25	39	.44	.41	.28	44	.42	.52	.32	36
	T													.45	.42	.37	106	.51	.50	.37	109	.55	.63	.49	94
11	H																	.58	.57	.54	47				1
	M																	.56	.52	.42	35				0
	L																	.40	.42	.33	39				2
	T																	.52	.51	.44	121				
12	H																					.64	.61	.57	49
	M																					.49	.53	.39	42
	L																					.38	.38	.27	32
	T																					.52	.52	.43	123

Table 14. Verbal B Problem Standard Deviations for Multivariate Analyses
Total Age Groups Divided into High, Medium, and Low I.Q. Groups

Age at First Administration		Age at Different Administrations																							
		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H								0	.22	.30	.43	28												
	M								2	.25	.35	.31	13												
	L								5	.29	.32	.22	12												
	T					.23	.43	.30	7	.26	.33	.37	53												
8	H								0	.27	.29	.29	26	.27	.31	.33	27								
	M								1	.32	.25	.23	29	.29	.37	.25	31								
	L								1	.26	.29	.26	19	.41	.42	.34	25								
	T									.31	.29	.29	74	.34	.38	.32	83								
9	H									.31	.34	.34	27	.29	.32	.36	29	.29	.30	.30	30				
	M									.22	.27	.37	20	.38	.32	.33	30	.30	.39	.38	28				
	L									.22	.33	.38	14	.30	.30	.32	29	.35	.36	.41	26				
	T									.29	.32	.36	61	.33	.33	.35	88	.31	.37	.37	84				
10	H													.40	.37	.33	41	.25	.30	.31	41	.28	.33	.36	37
	M													.32	.34	.32	26	.28	.26	.23	25	.34	.27	.30	21
	L													.28	.36	.31	39	.38	.34	.30	44	.31	.31	.32	36
	T													.35	.37	.34	106	.33	.33	.32	109	.33	.33	.37	94
11	H																	.31	.33	.34	47				1
	M																	.24	.27	.33	35				0
	L																	.31	.37	.30	39				2
	T																	.30	.33	.34	121				
12	H																					.29	.28	.32	49
	M																					.32	.26	.38	42
	L																					.27	.28	.36	32
	T																					.31	.29	.37	123

Table 15. Picture Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at First Administration		Age at Different Administrations: 7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H	.53	.46	.47	26	.51	.57	.37	23	.81	.53	.63	6												
	M	.47	.32	.32	37	.55	.70	.33	35	.63	.69	.59	15												
	L	.54	.33	.32	34	.45	.43	.20	33	.61	.67	.56	18												
8	H					.51	.53	.55	32	.07	.07	-.05	4				0								
	M					.50	.36	.39	28	.31	-.02	-.04	3				0								
	L					.46	.27	.36	35	.59	.48	.30	12				1								
9	H									.63	.57	.57	10				1				0				
	M									.54	.31	.48	14				0				0				
	L									.47	.31	.37	23				0				0				
10	H													.81	.66	.64	3				0				
	M													.73	.70	.68	8				0				
	L													.59	.61	.68	23				0				
11	H																				1				
	M																				3				
	L																				7				
12	H																								
	M																								
	L																								

Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H	.45	.31	.41	26	.25	.29	.42	23	.17	.27	.32	6												
	M	.36	.42	.47	37	.27	.33	.38	35	.24	.23	.30	15												
	L	.29	.35	.43	34	.30	.41	.39	33	.18	.21	.23	18												
8	H					.27	.36	.39	32	.25	.45	.43	4				0								
	M					.37	.43	.42	28	.09	.33	.28	3				0								
	L					.34	.37	.46	35	.26	.30	.39	12				1								
9	H									.30	.39	.35	10				1				0				
	M									.23	.40	.37	14				0				0				
	L									.27	.31	.45	23				0				0				
10	H													.17	.13	.26	3				0				0
	M													.24	.41	.41	8				0				0
	L													.31	.36	.30	23				0				0
11	H																				1				0
	M																				3				0
	L																				7				0
12	H																								3
	M																								2
	L																								4

Table 17. Verbal A Problem Means for Multivariate Analyses
Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				1				6	.70	.74	.58	26												
	M				0				2	.62	.70	.32	23												
	L				0				0	.64	.61	.34	24												
8	H								3	.55	.62	.57	29	.72	.74	.62	30								
	M								1	.57	.60	.30	23	.60	.69	.49	23								
	L								1	.61	.56	.30	24	.63	.55	.48	30								
9	H									.71	.70	.77	30	.64	.57	.49	33	.85	.78	.70	31				
	M									.57	.59	.60	28	.58	.49	.49	31	.81	.76	.66	30				
	L									.55	.54	.56	16	.53	.60	.37	25	.64	.68	.42	25				
10	H													.82	.75	.73	31	.66	.79	.72	29	.80	.83	.75	27
	M													.62	.60	.67	36	.68	.72	.66	34	.74	.78	.76	29
	L													.51	.55	.42	47	.49	.53	.39	47	.71	.67	.50	38
11	H																	.79	.80	.79	44				0
	M																	.73	.69	.61	47				2
	L																	.62	.55	.53	36				1
12	H																					.70	.77	.73	35
	M																					.69	.78	.71	49
	L																					.67	.65	.56	41

Table 18. Verbal A Problem Standard Deviations for Multivariate Analyses Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				1				6	.22	.22	.31	26												
	M				0				2	.25	.24	.34	23												
	L				0				0	.30	.31	.44	24												
8	H								3	.33	.28	.42	29	.23	.25	.31	30								
	M								1	.29	.26	.32	23	.38	.27	.37	23								
	L								1	.32	.33	.38	24	.42	.35	.35	30								
9	H									.29	.28	.24	30	.26	.29	.43	33	.14	.24	.21	31				
	M									.21	.26	.25	28	.37	.30	.30	31	.16	.23	.32	30				
	L									.32	.44	.24	16	.31	.26	.40	25	.35	.34	.42	25				
10	H													.13	.21	.22	31	.22	.18	.34	29	.22	.23	.22	27
	M													.28	.31	.39	36	.26	.20	.31	34	.31	.32	.29	29
	L													.34	.33	.35	47	.38	.34	.39	47	.23	.30	.29	38
11	H																	.23	.18	.22	44				0
	M																	.16	.27	.30	47				2
	L																	.28	.30	.35	36				1
12	H																					.31	.24	.26	35
	M																					.33	.16	.27	49
	L																					.30	.25	.26	41

Table 19. Verbal B Problem Means for Multivariate Analyses, Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at First Administration		Age at Different Administrations: 7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				0				5	.54	.46	.33	22												
	M				0				2	.43	.39	.07	15												
	L				0				0	.55	.45	.42	16												
8	H								1	.47	.41	.36	27	.57	.50	.41	30								
	M								1	.37	.23	.26	24	.43	.34	.37	22								
	L								0	.33	.31	.31	23	.42	.39	.24	31								
9	H									.72	.69	.59	27	.49	.56	.41	32	.68	.65	.52	30				
	M									.45	.48	.43	26	.53	.38	.34	31	.65	.57	.53	30				
	L									.41	.51	.19	8	.26	.41	.24	25	.45	.43	.30	24				
10	H													.60	.54	.51	29	.66	.64	.52	29	.71	.81	.60	27
	M													.51	.44	.38	32	.51	.55	.45	34	.55	.67	.61	29
	L													.30	.32	.27	45	.42	.39	.21	46	.44	.46	.32	37
11	H																	.63	.60	.54	42				0
	M																	.46	.48	.40	45				2
	L																	.44	.43	.38	34				1
12	H																					.57	.59	.52	34
	M																					.58	.56	.47	48
	L																					.41	.43	.30	41

Table 20. Verbal B Problem Standard Deviation for Multivariate Analyses Total Age Groups Divided into High, Medium, and Low G.P.A. Groups

Age at Different Administrations

Age at First Administration		7				8				9				10				11				12			
		31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n	31	33	35	n
7	H				1				5	.23	.33	.35	22												
	M				0				2	.27	.30	.30	15												
	L				0				0	.27	.35	.37	16												
8	H								1	.30	.30	.25	27	.31	.33	.31	30								
	M								1	.28	.28	.31	24	.31	.38	.27	22								
	L								0	.34	.24	.31	23	.38	.41	.33	31								
9	H									.19	.26	.32	27	.34	.29	.31	32	.23	.33	.34	30				
	M									.29	.36	.39	26	.27	.33	.34	31	.26	.33	.32	30				
	L									.28	.23	.20	8	.33	.32	.37	25	.39	.41	.41	24				
10	H													.38	.31	.29	29	.24	.30	.32	29	.26	.20	.39	27
	M													.34	.35	.37	32	.31	.32	.29	34	.36	.35	.23	29
	L													.28	.40	.30	45	.35	.32	.28	46	.29	.31	.36	37
11	H																	.26	.29	.35	42				0
	M																	.28	.34	.34	45				2
	L																	.34	.35	.29	34				1
12	H																					.31	.29	.37	34
	M																					.30	.28	.31	48
	L																					.31	.26	.40	41

Table 21. Multivariate Analysis Results: Comparisons of Different Age Groups

Lang.	Yr. of Test	F (Levels) Scheffe'	Individual Comparisons	Par.	F (Struct.) Scheffe'	31-33	31-35	33-35	Tot.	Number of Subjects for Age 12	11	10	9	8	7
P	1	.01	(7-10)*** (8-10)** (9-10)* (7,8,9-10,11,12)***	Yes	.01	.01	.05		293	9	11	34	47	95	97
	2	.05	(7-8)**	Yes	.01		.01	.01	110					19	91
	3														39
VA	1	.05	(10-12)* (10-11,12)*	Yes	.05		.10	.05	445	125	127	114	74	5	
	2	.10	(8,9-10)* (9-10)*	Yes	.01		.01	.01	283			110	89	75	8
	3	.01	(7-9)** (7-10)*** (7,8-9,10)*** (8-9)* (8-10)*	Yes	.01		.01	.01	336			94	86	83	73
VB	1	.05	(9-10)** (10-11,12)*	Yes	.01		.01	.01	411	123	121	106	61		
	2	.05	(8-10)**	Yes	.01		.01	.01	278			109	88	74	7
	3	.01	(8-10)*** (8-9)** (7-10)** (7-9)* (7,8-9,10)***	Yes	.01		.01	.01	313			93	84	83	53
	Str.				Yr. of Test	1-2	1-3	2-3							
P															39
VA	31		non-sig.	Yes	.01		.01	.01	177			92	76	9	
	33		non-sig.	Yes	.01		.01	.01	165			91	69	5	
	35		non-sig.	Yes	.01		.05	.01	152			88	60	4	
VB	31	.10	(9-10)*	Yes	.01		.01	.05	136			80	56		
	33	.10	(9-10)*	Yes	.01		.01	.01	132			79	53		
	35	.10	(9-10)*	Yes	.01		.01	.01	130			78	52		

Table 22. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups Over the Total Sample

Lang.	Yr. of Test	F (Levels)	Scheffe' H-M	H-L	M-L	Par.	F (Struct.)	Scheffe' 31-33	31-35	33-35	Number of Subjects Tot.	H	M	L	I. Q. H	M	L
P	1	.05				Yes	.01				293	87	89	117	122	108	95
				.10				.01	.01								
	2	non-sig.				Yes	.01				111	35	30	46	124	108	96
									.01	.01							
	3	.10				Yes	non-sig.				40	9	12	19	124	108	97
				.10													
VA	1	.01				Yes	.05				446	171	138	137	121	109	94
			.01	.01	.01				.10	.05							
	2	.01				Yes	.01				286	103	87	96	122	108	95
			.05	.01					.01	.01							
	3	.01				Yes	.01				340	129	99	112	123	108	95
			.05	.01	.01				.01	.01							
VB	1	.01				Yes	.01				413	165	124	124	121	108	94
			.01	.01	.01				.01	.01							
	2	.01				Yes	.01				281	102	86	93	122	108	95
			.01	.01					.01	.01							
	3	.01				Yes	.01				317	123	93	101	122	108	95
			.10	.01	.05				.01	.01							
	Str.						Yr. of Test	1-2	1-3	2-3							
P	31	non-sig.				Yes	.05				39	9	11	19	124	108	97
										.10							
	33	non-sig.				Yes	.01				38	9	10	19	124	107	97
		.10							.01	.01							
	35			.10		Yes	.01				34	8	8	18	122	108	97
VA		.10							.01	.01							
	31	.01				Yes	.01				180	67	54	59	122	108	94
				.01	.01				.01	.01							
	33	.01				Yes	.01				168	66	48	54	122	108	94
			.01	.01					.01	.01							
	35	.01				Yes	.01				155	61	46	48	122	108	95
VB			.01	.01					.01	.01							
	31	.01				Yes	.01				142	58	42	42	123	108	94
			.05	.01					.01	.01							
	33	.01				Yes	.01				139	58	41	40	123	108	95
				.01					.01	.01							
	35	.01				Yes	.01				135	58	37	40	123	108	94
			.01	.01					.01	.01							

Table 23. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups for 7 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Number of Subjects Tot.	H	M	L	I. Q. H	M	L
P	1	non-sig.				Yes	.01	.01	.01		97	39	26	32	124	108	95
	2	.10		.10		Yes	.01		.01	.01	91	33	27	31	124	108	95
	3	.10		.10		Yes	non-sig.				39	9	12	18	124	108	97
VA	1										1	1	0	0			
	2										8	6	2	0			
	3	.01		.01		Yes	.01		.01	.01	73	34	19	20	124	108	95
VB	1										0	0	0	0			
	2										7	5	2	0			
	3	.10		.10		Yes	.01		.01	.05	53	28	13	12	124	109	95
	Struct.						Yr. of Test	1-2	1-3	2-3							
P	31	non-sig.				Yes	.05			1,2,3	38	9	11	18	124	108	97
	33	non-sig.				Yes	.01		.01	.05	37	9	10	18	124	107	97
	35	.10		.10		Yes	.01		.01	.01	33	8	8	17	123	108	96
VA												0	0	0			
VB												0	0	0			

Table 24. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups for 8 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Number of Subjects Tot.	H	M	L	I. Q. H	I. Q. M	I. Q. L
P	1	.10		.10		YES	.10	.10			95	29	34	32	122	109	98
	2										19	2	3	14			
	3										1	0	0	1			
VA	1										5	1	4	0			
	2	.01		.01	.10	Yes	.01		.01	.01	76	26	30	20	123	108	98
	3	.01	.01	.01		Yes	.01		.05	.01	83	27	30	26	122	109	98
VB	1											1	1	0			
	2	.01	.10	.01	.10	Yes	.05		.10		74	26	29	19	123	108	98
	3	.01	.10	.01		Yes	.01		.01		83	27	31	25	122	109	98
	Str						Yr. of Test	1-2	1-3	2-3							
P												0	0	1			
VA	31											2	5	2			
	33											2	3	0			
	35											1	3	0			
VB	31											1	2	0			
	33											1	2	1			
	35											1	1	0			

Table 25. Multivariate Analysis Results:
Comparisons of High, Medium and Low I.Q. Groups for 9 Year Olds

Lang.	Yr. of Test	F (Levels)			Par.	F (Struct.)			Number of Subjects				I. Q.		
		Scheffe'				Scheffe'									
		H-M	H-L	M-L		31-33	31-35	33-35	Tot.	H	M	L	H	M	L
P	1	non.sig.			Yes	.10			47	10	14	23	120	106	95
						.10									
	2								1	0	0	1			
	3								0	0	0	0			
VA	1	.10			Yes	Non-sig.			74	30	25	19	123	108	98
			.10												
	2	non-sig.			Yes	.01			89	29	30	30	123	108	97
							.05	.05							
	3	.05			Yes	.01			86	30	28	28	123	108	97
			.05	.10			.01	.01							
VB	1	non-sig			Yes	.10			61	27	20	14	124	108	98
							.10	.10							
	2	.10			Yes	.01			88	29	30	29	124	108	98
			.10				.10	.01							
	3	non-sig			Yes	.01			84	30	28	26	123	108	97
							.01	.05							
	Str.					Yr. of Test									
						1-2	1-3	2-3							
P										0	0	0			
VA	31	non-sig			Yes	.01			76	27	27	22	123	108	98
							.01	.01							
	33	1st 2nd 3rd .05			No	L.10 M.01 H.01			69	26	23	20	123	109	98
	35	.01			Yes	.01			60	22	21	17	124	109	98
			.01			.0505		.01							
VB	31	1st .01 2nd 3rd .10			No	L.01 M M.10			56	21	21	14	124	109	99
	33	non-sig			Yes	.01			53	21	20	12	124	109	99
								.01							
	35	non-sig			Yes	.05			52	21	19	12	124	109	99
							.09	.10							

Table 26. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups for 10 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe'			Par.	F (Struct.)	Scheffe			Number of Subjects				I.Q.		
			H-M	H-L	M-L			31-33	31-35	33-35	Tot.	H	M	L	H	M	L
P	1	non-sig.				Yes	non-sig				34	6	9	19	117	107	90
	2										0	0	0	0			
	3										0	0	0	0			
VA	1	.01		.01	.10	Yes	non-sig.				114	42	29	43	122	108	92
	2	.01		.01	.10	Yes	.05			.05	111	41	25	45	122	108	90
	3	.01	.01	.01		Yes	.01		.01	.01	95	37	22	36	121	108	89
VB	1	.01	.10	.01		Yes	non-sig				106	41	26	39	122	108	92
	2	.01	.01	.01		Yes	.01		.01	.01	110	41	25	44	121	108	90
	3	.01	.10	.01		Yes	.01	.05		.01	94	37	21	36	121	108	89
	Str.							1-2	1-3	2-3							
P											0	0	0	0			
VA	31	.01		.01	.01	Yes	.01		.01	.01	93	37	22	34	121	108	89
	33	.01	.05	.01		Yes	.01		.01	.10	92	37	22	33	121	108	89
	35	.01	.01	.01		Yes	.05		.05	.10	88	37	22	29	121	108	92
VB	31	.01	.05	.01		Yes	non-sig.				80	35	19	26	121	108	92
	33	.01	.05	.01		Yes	.01		.01	.01	79	35	19	25	121	108	93
	35	.01	.01	.01		Yes	.01		.05	.01	78	35	17	26	121	108	92

Table 27. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups for 11 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Number of Subjects Tot.	H	M	L	I. Q. H	M	L
P	1										11	2	1	8			
	2										0	0	0	0			
	3																
VA	1	.01		.01	.05	Yes	.05		.05		127	48	36	43	120	109	93
	2																
	3																
VB	1	.01		.01		Yes	.05		.10		121	47	35	39	120	109	93
	2																
	3																

Table 28. Multivariate Analysis Results:
Comparisons of High, Medium, and Low I.Q. Groups for 12 Year Olds

Lang.	Yr. of Test	F (Levels)				F (Struct.)			Number of Subjects				I.Q.		
		Scheffe				Scheffe									
		H-M	H-L	M-L		31-33	31-35	33-35	Tot.	H	M	L	H	M	L
P									0	0	0	0			
VA	1	.01			Yes	.05			125	49	44	32	120	109	95
		.01	.01					.05							
VB	1	.01			Yes	.01			123	49	42	32	120	109	95
		.05	.01	.10			.05	.01							

Table 29. Multivariate Analysis Results:

Comparisons of High, Medium, and Low G.P.A. Groups for 7 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Tot.	Number of Subjects H	M	L	G.P.A. H	M	L
P	1	non-sig.				Yes	.01	.01	.01		97	26	37	34	3.81	2.98	1.91
	2	.05			.05	Yes	.01		.01	.01	91	23	35	33	3.82	2.99	1.81
	3	non-sig.				Yes	non-sig.				39	6	15	18	4.37	2.95	1.92
VA	1											1	0	0			
	2											6	2	0			
	3	.05		.10		Yes	.01		.01	.01	73	26	23	24	3.92	3.01	1.90
VB	1											0	0	0			
	2											5	2	0			
	3	non-sig.				Yes	.01		.01	.05	53	22	15	16	3.83	3.04	1.78
	Str.						Yr. of Test	1-2	1-3	2-3							
P	31	not sig.				Yes	.05			.10	38	5	15	18	4.48	2.95	1.92
	33	not sig.				Yes	.01		.01	.01	37	5	15	17	4.45	2.95	1.89
	35										33	2	15	16			
VA	31											0	0	0			
VB	33											0	0	0			

Table 30. Multivariate Analysis Results:

Comparisons of High, Medium, and Low G.P.A. Groups for 8 Year Olds

<table>
<tr><th rowspan="3">Lang.</th><th rowspan="3">Yr. of Test</th><th colspan="3">F (Levels)</th><th rowspan="3">Par.</th><th colspan="3">F (Struct.)</th><th rowspan="3">Tot.</th><th colspan="3">Number of Subjects</th><th colspan="3">G.P.A.</th></tr>
<tr><th colspan="3">Scheffe</th><th colspan="3">Scheffe</th><th colspan="3"></th><th colspan="3"></th></tr>
<tr><th>H-M</th><th>H-L</th><th>M-L</th><th>31-33</th><th>31-35</th><th>33-35</th><th>H</th><th>M</th><th>L</th><th>H</th><th>M</th><th>L</th></tr>
<tr><td>P</td><td>1</td><td colspan="3">.10</td><td>Yes</td><td colspan="3">.10</td><td>95</td><td>32</td><td>28</td><td>35</td><td>4.65</td><td>3.49</td><td>1.98</td></tr>
<tr><td></td><td></td><td></td><td>.10</td><td></td><td></td><td>.10</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>2</td><td colspan="3">.01</td><td>Yes</td><td colspan="3">not sig.</td><td>19</td><td>4</td><td>3</td><td>12</td><td>4.50</td><td>3.30</td><td>2.24</td></tr>
<tr><td></td><td></td><td></td><td>.05</td><td>.10</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>0</td><td>0</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>VA</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>5</td><td>3</td><td>1</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td></td><td>2</td><td colspan="3">not sig.</td><td>Yes</td><td colspan="3">.01</td><td>76</td><td>29</td><td>23</td><td>24</td><td>4.68</td><td>3.53</td><td>1.70</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>.01</td><td>.01</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>3</td><td colspan="3">not sig.</td><td>Yes</td><td colspan="3">.01</td><td>83</td><td>30</td><td>23</td><td>30</td><td>4.68</td><td>5.53</td><td>1.85</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>.01</td><td>.01</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>VB</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>1</td><td>1</td><td>0</td><td></td><td></td><td></td></tr>
<tr><td></td><td>2</td><td colspan="3">not sig.</td><td>Yes</td><td colspan="3">.10</td><td>74</td><td>27</td><td>24</td><td>23</td><td>4.70</td><td>3.54</td><td>1.78</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td>.10</td><td>.10</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>3</td><td colspan="3">not sig.</td><td>Yes</td><td colspan="3">.01</td><td>83</td><td>30</td><td>22</td><td>31</td><td>4.68</td><td>3.54</td><td>1.87</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>.01</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="3"></td><td></td><td colspan="3">Yr. of Test</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>Str.</td><td></td><td></td><td></td><td></td><td>1-2</td><td>1-3</td><td>2-3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>P</td><td>31</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>0</td><td>0</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td></td><td>33</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>0</td><td>0</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td></td><td>35</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td>0</td><td>0</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>VA</td><td>31</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>9</td><td>3</td><td>1</td><td>5</td><td></td><td></td><td></td></tr>
<tr><td></td><td>33</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>5</td><td>2</td><td>1</td><td>2</td><td></td><td></td><td></td></tr>
<tr><td></td><td>35</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>4</td><td>2</td><td>1</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>VB</td><td>31</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>3</td><td>1</td><td>1</td><td>1</td><td></td><td></td><td></td></tr>
<tr><td></td><td>33</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>4</td><td>1</td><td>1</td><td>2</td><td></td><td></td><td></td></tr>
<tr><td></td><td>35</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>1</td><td>1</td><td>0</td><td></td><td></td><td></td></tr>
</table>

Table 31. Multivariate Analysis Results:

Comparisons of High, Medium, and Low G.P.A. Groups for 9 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Number of Subjects Tot.	H	M	L	G.P.A. H	M	L
P	1	non-sig.				Yes	.05				47	23	14	10	5.44	4.05	2.71
								.05									
	2											1	0	0			
	3											0	0	0			
VA	1	.05				Yes	non-sig.				74	30	28	16	5.72	4.13	2.03
				.05													
	2	non-sig.				Yes	.05				89	33	31	25	5.63	4.08	2.12
									.05	.05							
	3	.01				Yes	.01				86	31	30	25	5.64	4.07	2.08
				.01	.05				.01	.01							
VB	1	.01				Yes	.05				61	27	26	8	5.72	4.11	1.28
			.05		.05				.10	.10							
	2	.05				Yes	.01				85	32	31	25	5.64	4.08	2.12
				.05					.05	.01							
	3	.05				Yes	.01				84	30	30	24	5.64	4.07	2.03
				.10	.10				.01	.01							
	Str.							Yr. of Test 1-2	1-3	2-3							
P	31											0	0	0			
	33											0	0	0			
	35											0	0	0			
VA	31	.05				Yes	.01				76	29	27	20	5.66	4.08	2.10
				.05					.01	.01							
	33	non-sig.				Yes	.01				69	26	25	18	5.66	4.10	2.08
									.01	.01							
	35	.10				Yes	.01				60	25	21	14	5.69	4.08	1.86
				.10				.05		.01							
VB	31	.01				Yes	.01				56	23	21	12	5.75	4.08	1.77
				.01	.05				.05	.01							
	33	non-sig.				Yes	.01				53	23	21	9	5.75	4.08	1.49
										.01							
	35	.10				Yes	.05				52	23	21	8	5.75	4.08	1.28
									.10	.05							

Table 32. Multivariate Analysis Results: Comparisons of High, Medium, and Low G.P.A. Groups for 10 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Tot.	Number of Subjects H	Number of Subjects M	Number of Subjects L	G.P.A. H	G.P.A. M	G.P.A. L
	1	non-sig.				Yes	non-sig.				34	3	8	23	4.77	4.09	2.11
P	2											0	0	0			
	3											0	0	0			
	1	.01	.10	.01	.05	Yes	non-sig.				114	31	36	47	5.47	4.14	2.10
VA	2	.01		.01	.01	Yes	.05			.05	110	29	34	47	5.47	4.12	2.21
	3	.01		.05	.10	Yes	.01		.01	.01	94	27	29	38	5.41	4.13	2.22
	1	.01		.01	.10	Yes	non-sig.				106	29	32	45	5.51	4.11	2.07
VB	2	.01		.01	.05	Yes	.01		.01	.01	109	29	34	46	5.47	4.12	2.20
	3	.01		.01	.05	Yes	.01	.05		.01	93	27	29	37	5.41	4.13	2.20
	Str.						Yr. of Test	1-2	1-3	2-3							
	31										0						
P	33										0						
	35										0						
	31	.01		.01	.01	Yes	.01		.01	.01	92	27	28	37	5.41	4.14	2.19
VA	33	.01		.01	.10	Yes	.01		.01	.10	91	27	29	35	5.41	4.13	2.18
	35	.01		.01	.01	Yes	.05		.05	.05	88	26	29	33	5.44	4.13	2.16
	31	.01	.10	.01		Yes	non-sig.				80	24	24	32	5.50	4.10	2.16
VB	33	.01	.10	.01	.10	Yes	.01		.01	.01	79	24	24	31	5.50	4.11	2.14
	35	.01		.01	.05	Yes	.01		.05	.01	78	24	24	30	5.50	4.11	2.15

Table 33. Multivariate Analysis Results:
Comparisons of High, Medium, and Low G.P.A. Groups for 11 Year Olds

Lang.	Yrs. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Tot.	Number of Subjects H	Number of Subjects M	Number of Subjects L	G.P.A. H	G.P.A. M	G.P.A. L
P	1										11	1	3	7			
	2										0	0	0	0			
	3																
VA	1	.01	.05	.01	.05	Yes	.05		.05		125	44	47	36	5.30	4.10	2.64
	2										3	0	2	1			
	3																
VB	1	.01	.05	.01		Yes	.05			.10	121	42	45	34	5.30	4.10	2.57
	2										3	0	2	1			
	3																

Lang.	Str.		Par.	Yr. of Test 1-2	Yr. of Test 11-3	Yr. of Test 2-3	Tot.	H	M	L	G.P.A.
P	31						0	0	0	0	
	33						0	0	0	0	
	35						0	0	0	0	
VA	31						0	0	0	0	
	33						0	0	0	0	
	35						0	0	0	0	
VB	31						0	0	0	0	
	33						0	0	0	0	
	35										

Table 34. Multivariate Analysis Results:
Comparisons of High, Medium, and Low G.P.A. Groups for 12 Year Olds

Lang.	Yr. of Test	F (Levels)	Scheffe H-M	Scheffe H-L	Scheffe M-L	Par.	F (Struct.)	Scheffe 31-33	Scheffe 31-35	Scheffe 33-35	Tot.	Number of Subjects H	Number of Subjects M	Number of Subjects L	G.P.A. H	G.P.A. M	G.P.A. L
P	1										9	3	2	4			
	2																
	3																
VA	1	.05	.05	.05		Yes	.05			.05	125	35	49	41	5.31	4.08	2.78
	2																
	3																
VB	1	.01		.01	.05	Yes	.01		.05	.01	123	34	48	41	5.31	4.08	2.78
	2																
	3																

Lang.	Str.			Yr. of Test 1-2	Yr. of Test 1-3	Yr. of Test 2-3	Tot.	H	M	L			
P	31						0	0	0	0			
	33						0	0	0	0			
	35						0	0	0	0			
VA	31						0	0	0	0			
	33						0	0	0	0			
	35						0	0	0	0			
VB	31						0	0	0	0			
	33						0	0	0	0			
	32						0	0	0	0			

Conclusions and Recommendations

In attempting to interpret and summarize the results of this study, it is necessary to highlight some of the unique features of this research. These refer to theoretical and methodological considerations as well as to the experimental procedures used.

With regard to theoretical considerations we attempted to separate the concepts of logical sttucture and language and subsequently to develop a procedure that would permit their experimental identification. This resulted in the preparation of instruments (problems) strictly defined, so that in terms of these two components, language and logical structure, a classification of problems could be made. Isomorphic problems were defined as resulting from mapping the same logical structures using different manners of expression.

If the languages used are different (P, VA, VB) a series of isomorphic problems could be of the type 31P, 31VA, 31VB, and similarly for other structures. This corresponds to those problems described in cell III of Table 4. If the structure and the language are the same but the words or symbols used are different, the series would be represented by the sequence (for instance when using the VA language using different words) $31VA_1$, $31VA_2$, $31VA_3$ in which 1, 2, 3, refer to administration but also to the concrete manner of presentation of the problem. These correspond to the problems described in cell I of Table 4. It could be said that problems of the 1st type are, as it were, translations of the same structure in different languages, while problems of the 2nd type are synonymous.

It follows that the comparison of performance in problems like those of cell III in Table 4 should give information as to how language affects the subject's tactics for a specified structure. The comparison of problems like those corresponding to cell I in Table 4 will give indication as to how the use of a given language changes through successive years.

If a subject solves at least one of all the isomorphic problems administered to him, then we are entitled to conclude that he can deal with the corresponding logical structure. By using isomorphic problems it is then possible to establish thresholds for the different structures. If a subject can solve a problem with structure 31, then he is at or above the threshold corresponding to that structure. On the other hand if he can solve problem 31P, but not problem 31VA, then we could say that though he is at or above threshold for structure 31, he can not yet operate successfully with the VA language. Of course this can only refer to the special words used in presenting problem 31VA since a synonymous problem may be solved. In this research, synonymous problems were used to study the effects of successive administrations.

If a subject can operate satisfactorily with a given language, then comparison of problems presented in the same language but built around different structures should give information concerning his understanding of the logical structures involved. Problems of this type correspond to those of cell II in Table 4, for instance 31P, 33P, 35P.

The different comparisons implied in the previous description were performed in this research. Their cross examination provides a check on each specific finding so that we could test thresholds for logical structure through performances in isomorphic problems (translations and synonymous problems) as well as using one language and different structures. Similarly, language thresholds can be checked by examining when a subject can solve a problem in a given language. For instance if a subject can solve problem 33VA, but not 33VB, then we will say that he is below threshold for language VB.

The comparisons performed in this research are given in Tables 21 to 34. They were performed independently using several subclassifications of subjects, namely chronological age, I.Q., and G.P.A. The similarities or differences obtained should then be of interest to appraise the influence of these variables in problem solving performances. Further, changes in successive performances (first, second and third administration) were performed for problems with the same language and same structure.

As explained in the Introduction to this study, we examine subject's tactics, a tactic being defined experimentally by the sequence of questions that a subject asks. It is assumed that tactics are images of the subject's processes (the correspondence being either of the one-one type or of the many-one type). Usually in studying cognitive processes, experimenters rely on the final answers given to problems, and as our experience has shown, it is risky to infer processes from final answers. On the other hand a thinking aloud approach, or other similar experimental situations make it difficult to compare results, mostly when different researchers are involved.

Therefore the problems used in this study are a compromise between a situation that leaves both subject and experimenter a considerable degree of freedom and a situation in which the interpretation of the responses may be widely divergent from what subjects really do. As stated previously our main interest is in the study of tactics, and these are made observable through the analyses of the questions that the subject asks.

This brings into the foreground methodological problems that do not weigh so heavily in other types of tests. One of them refers to the order in which questions are asked. In years past we scored problems in terms of indexes developed from the performances of selected samples. But as discussed in the Introduction, the same performance may obtain different evaluations according to the normative sample used. On the other hand it might be desirable to score subjects' tactics regardless of the sample to which he belongs. Thus we have attempted through the years to develop a scoring procedure such that it will depend exclusively on the properties of the problem used and on the subject's ability, as expressed in his tactic.

In the section on Method this approach is described in detail. Basically scores depend on how the subject handles the logical structure of the problem. The ideal tactic is the one that follows the logical structure closely and most parsimoniously without reversals or redundancies. The result is that scores can be assigned to specific tactics based exclusively on how a subject approaches the logical structure of the problem. Thus isomorphic problems should obtain the same pulling out score regardless of the language used, provided the tactic is the same in all of them. Otherwise the observed pulling out score will vary, and these changes in score should reflect the "difficulty" of the problem as presented. The concept of "intrinsic" difficulty of a problem relates to its logical structure. The concept of extrinsic difficulty depends on the language used. If two isomorphic problems give two different pulling out scores in the same subject, then we are entitled to say that the difference is due to the language used. Thus for a given structure the intrinsic difficulty is constant no matter which language is used.

The pulling out scoring method used in this study is an improvement on the previous pulling out scoring methods that were used in the past. The procedure used here allows the comparison of scores for different structures along a continuum that varies from -1.00 to 1.00. In all likelihood the method that we use here goes beyond a simple rank ordering of values. If this is demonstrated then logical structures can be placed in terms of their difficulty within the interval indicated above. It is intuitively clear by analyzing the results of this research, that this condition has probably been fulfilled.

It should also be noticed that the scoring method used here considers as a very important variable the relation of procedence among questions. In a temporal context this means that logical structures can be examined psychologically as temporarily ordered events.

The fact that we are not using sampling procedures to evaluate tactics does not mean that statistical operations can not be performed to define samples and evaluate subjects. As a matter of fact the type of score that is used here is similar to the one used in those measurements for which a ruler has been defined. Applying such ruler to an object, results (scores) can be used to classify the observations into different subsets and the corresponding statistics can be performed. In the present research the properties of the ruler used are defined in terms of logical concepts. Basically these refer to the properties of ordered sets.

The statistical analyses performed could have been more extensive and other methods could have been used. The vast amount of data accumulated required a procedure that would allow dealing with several variables simultaneously. Further since the same subjects were tested at least three times, covariances play a definite role in the interpretation of the results. Therefore early in our research it was decided to use a multivariate analysis appraoch, more specifically profile analysis for several independent groups. The independent groups refer to the classification of subjects in terms of chronological age, or I.Q., or G.P.A. averages. The covariance matrices refer to successive administration of the same problems or of isomorphic problems or of problems varying in both language and structure as the case may be. An important facet of this study relates to the study of specified contrasts (age groups, I.Q. levels, G.P.A. levels, etc.). The Scheffé technique was used for determining these simultaneous confidence intervals.

For completeness of information we have included summary statistics for all the problems used in this study. These appear in Table 5. Other analyses could have been performed to discuss our data. Nevertheless because of the purposes of the study and the straight forwardness of the multivariate formulation, the procedures used were thought to fit our aims best.

The experimental procedure used with regard to the administration of the problems implies testing simultaneously for both structure and language thresholds. Those subjects who could not solve problems in P language were not administered VA or VB problems. Similarly subjects who at later ages could not solve VA problems were administered P problems. On the basis of our previous research an increasing difficulty level was postulated progressing from P to VA to VB for languages and from 31 to 33 to 35 for structures. The results of this study seem to indicate that our "guess" was essentially correct, at least within the age groups considered in this investigation. But by doing this, and by not using those subjects that did not meet the minimum requirements specified in the section on Method, the sample used was at or above threshold at least for the simple structure (31) and for the P language. Nevertheless the differential thresholds between structures and languages could be tested. As shown in the section on Results, they were in both cases significant (Table 21). It seems quite clear that between 9 and 10 years of age there is a significant change in level of performance when using P language. A similar change is observed approximately a year later with VA and VB languages. These results seem to indicate that for the age groups used in this study, two relatively homogeneous subsamples can be defined with reference to language P; one including ages 7, 8 and 9, and another including ages 10, 11 and 12. With the VA and VB languages a similar phenomenon is observed approximately one year later. Further it was suggested that with P language differentiation of problems tends to become less clear at later ages, though with the VA and VB languages the opposite seems to be true. The total sample, including all ages has shown that structures are differentiated in all the languages used. Also that contrasts based on administration session, for problems with same structure and language, show that the scores increase from first and second administration to third administration. Thus throughout the years there is a progressive improvement in successive performances, when VA and VB languages are used. Nevertheless the comparisons between age levels for problems in the same language and same structure through the three years of testing do not give significantly different profiles. That is, while comparison of age levels is made on the basis of problems with same language but different structures (keeping administration constant) the age levels are clearly differentiated but not so when the problems have the same language and the same structure.

The educational and psychological implications of this observation are of interest, in the sense that the presentation of a given concept may be more readily understood in one language than in another. That is given that a subject or a student has reached the threshold that corresponds to a given logical structure, then it should be a question of presenting it in the language that is most appropriate. A systematic analysis of thresholds of different logical structures should be conducted and recommendations concerning the most effective manner of presentation should be an integral part of such study. But in doing so the burden of the proof should be left on the student in the sense that the student's tactics, that is how they solve a problem, should tell us how each individual subject performs. This is obviously different from appraising a student's ability by noting how many problems he answered correctly or incorrectly. In this situation no tactic is involved, and possible detours, failures in reasoning, missing the crucial elements, simple adhering to a pattern, etc., cannot be discovered except in roundabout ways.

Once these thresholds have been determined it should be a relatively simple matter to find out which language or manner of presentation increases the possibility of discriminating among structures. Our results seem to indicate that the P language is better fitted to fulfill this purpose in the early years, but becomes less discriminating with increasing age. With the VA and VB languages discrimination gets better with advancing age.

Our results also suggest that average evaluations may risk ignoring the outstanding and the backward children as in the case of those 7 and 8 year olds that performed better than their peers. In this respect it is worth noticing that since our scoring system does not depend on sampling performances, it would be possible to place a subject in a specified category regardless of the age group to which he belongs. His score will depend exclusively on how he understands the logical structure of the problem.

With regard to profiles based on I.Q. levels (Table 22) comparing the performance of problems with same language but different structures in the same administration, the results indicate that different I.Q. levels give significantly different results. This is basically due to problems in VA and VB languages and much less in P problems. Structures are as well differentiated as they were in the case of age levels. However the comparison of I.Q. profiles using problems with the same language and same structure through the three years of testing shows highly significant results, with good differentiation between H-L and H-M contrasts. Also the third administration gives higher results than the first or the second ones. Pooling together this information and comparing it with the one obtained using an independent criterion for classification (age level) suggests some interesting possibilities. One of these is that I.Q. levels may depend quite heavily on the language used.

That is, inability to operate with a given language may show in reduced test scores, not because the subjects are unable to understand the structure of the problem but because they are less proficient in the use of the manner of expression in which the problem is presented. These observations verify previous findings listed in the Introductory section. The implication being that in studies concerning language deficiency, perceptual deficiency, etc., it is not enough to conclude that the cognitive thinking ability of a subject is low because of his failure to solve one or several problems. The questions should rather be, 1) does he understand the logical structure of the problems, 2) can he use the language efficiently. A negative answer to the first question is quite final, but a negative answer to the second one does not say much about the first. This, of course, has implication in intercultural studies. As a matter of fact, fluency in a language is not an indication that the subject can understand complex or even simple logical structures.

Comparisons of profiles in age or I.Q. levels with problems of the type $31VA_i$, $33VA_i$, $35VA_i$, where i is any fixed administration, were shown in general to be significantly different. The same type of comparisons of profiles defined by isomorphic problems, for instance $31VA_1$, $31VA_2$ and $31VA_3$ do not show significantly different levels when the profiles are based on age, though in terms of I.Q. levels they are significant. In all cases, whether using I.Q. or age, the comparisons between structures is significant as well as the contrasts involving administration (structure 31 and 33 are overall significantly different from structure 35 and the third administration gives better scores than the first or the second). Pooling together these results it seems reasonable to infer that in order to appraise a subject's ability it may be better to scan his performance in several problems with different structures than to administer to him several isomorphic problems. The question that becomes then focal is: should students be trained by practicing repeatedly with isomorphic problems, a few selected concepts, or should they be administered problems with different structures in a language that they can manipulate. Though it is true that educationally speaking, a certain minimum level of proficiency is desirable, the danger may reside in the possibility that by concentrating on this aspect the subjects with outstanding ability may be ignored. As it is well known this has considerable implications in education and in society in general. This seems to fall in line with some of the questions that some psychologists and educators are raising today with regard to some testing procedures commonly used to evaluate cognitive ability and thinking processes. It is then suggested that a large variety of problems with different logical complexity should be used to evaluate subjects' ability. But unfortunately if the structure of these problems is made in terms of how a group of subjects perform, then the risk is that the most discriminating problems (test items) will either be eliminated or would obtain a low weight in the final evaluation.

The existence of a transitional period with regard to the use of language seems to occur between 9 and 11 years of age. The pattern of significant contrasts before and after these ages seems to indicate that there is a shift from the use of P language towards the more efficient use of VA and VB languages at the later years. This is shown whether subjects are classified in terms of I.Q. levels or of G.P.A. scores. In terms of the latter, the profiles for the different G.P.A. levels in relation to the different contrasts examined, do not differ substantially from those obtained using I.Q. levels except for the fact that G.P.A. levels for the low and medium groups are more sharply differentiated than for the low and medium I.Q. levels. It appears that in grading children, teachers are more prone to differentiate poor from medium achievers, while I.Q. evaluations tend to differentiate better among subjects at higher ability levels. This is a finding that might have some interest with reference to grading systems. It should be remembered that when analyzing the distribution of G,P.A. averages it was found that the criterion for grading tended to be more strict for younger than for the older children.

The performance of subjects at specified ages was studied in terms of both I.Q. levels and G.P.A. scores. Comparisons were made between the findings obtained in both cases when discussing G.P.A. levels (Tables 23 to 34). The similarity of the results was noticed and specific differences were listed in the corresponding text. In general, significant contrasts became sharp or sharper with increasing age samples in VA and VB problems, while P problems seem to lose differentiating power at around 10 years of age. It is around 9 years of age that a transition occurs for the picture problems. In subsequent years when they are 10 and 11 years of age, a similar transition occurs for the verbal problems.

In summary it can be said that there is improvement of performance in the successive years, that subjects are able to differentiate between logical structures, and that the language used influences the performance in the problems.

References

Anderson, T. Introduction to Multivariate Statistical Analysis. New York, Wiley, 1958.

Attneave, F. Applications of Information Theory to Psychology. New York, Henry Holt and Company, 1959.

Bartlett, E. Thinking. New York, Basic Books, 1958.

Binet, A. L'etude Experimentale de L'intelligence. Paris, Scheiler, 1902.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A Study of Thinking. New York, Wiley, 1956.

Duncker, K. On problem solving. Psychol. Monographs, 1945, 58, No. 5 (whole No. 270).

Feller, W. An Introduction to Probability Theory and its Application. New York, Wiley, 1959.

Gunn, H. Appraisal of personality parameters in terms of processes, Loyola University, Chicago, Ill., 1959 (unpublished Ph.D. dissertation).

Haley, J. The effect of learning on performance in the test of diagnostic skills. Unpublished M.A. thesis, 1960, Loyola University, Chicago, Ill.

Hubbard, J. Programmed testing in the examination of the National Board of Medical Examiners, 1964 (presented at the proceedings of 1963 International Conference on Testing Problems, Educational Testing Service, New Jersey).

John, E., & Rimoldi, H. Sequential observation of complex reasoning. The American Psychologist, X, 1955, 470 (Abstract).

Kohler, W. The Mentality of Apes. New York, Harcourt, Brace, 1927.

Morrison, D. Multivariate Statistical Methods. New York, McGraw-Hill, 1967.

Rimoldi, H. A technique for the study of problem solving. Educational and Psychological Measurement, 1955, 15, 4, 450-461.

Rimoldi, H. Decision processes in mathematical thinking, 1967. (Report to National Science Foundation on a grant).

Rimoldi, H. Logical structure and languages in thinking processes. *International Journal of Psychology* (in press), 1970, Paris, France.

Rimoldi, H. Thinking and language. *Archives of General Psychiatry*, November, 1967, 17.

Rimoldi, H. On cognizing cognitive processes. (Read at Erikson Symposia on Growth and Development, October, 1969). Loyola Psychometric Laboratory publication no. 57 (1969), Loyola University, Chicago, Ill.

Rimoldi, H., Aghi, M., & Burger, G. Some effects of logical structure, language and age in problem solving in children. *Journal of Genetic Psychology*, 1968, 112, 127-143.

Rimoldi, H., Devane, J., & Haley, J. Characterization of processes. *Educational and Psychological Measurement*, 1961, 21, 383.

Rimoldi, H., Devane, J. Training in problem solving. Loyola Psychometric Laboratory publication no. 21 (1961), Loyola University, Chicago, Ill.

Rimoldi, H., Fogliatto, H., Haley, J., Reyes, I., Erdmann, J., & Zacharia, R. Training in problem solving (Cooperative Research Project no. 1449 (1962) supported by the Office of Education, U. S. Department of Health, Ecucation and Welfare) Loyola Psychometric Laboratory publication no. 27 (1962), Loyola University, Chicago, Ill.

Rimoldi, H., Fogliatto, H., Erdmann, J., Donnelly, M. Problem solving in high school and college students (Cooperative Research Project No. 2199 (1964) supported by the Office of Education, U. S. Department of Health, Education & Welfare) Loyola Psychometric Laboratory Publication No. 42 (1964), Loyola University, Chicago, Ill.

Rimoldi, H., Grib, T. Pattern analysis. *British Journal of Statistical Psychology*, Nov., 1960.

Rimoldi, H., Haley, T., Fogliatto, H., Erdmann, J. A program for the study of thinking. *Bulletin International Association of Applied Psychology*, France, 1964.

Rimoldi, H., Haley, J., Fogliatto, H. The test of diagnostic skills. Loyola Psychometric Laboratory Publication no. 25 (1962), Loyola University, Chicago, Ill.

Rimoldi, H., Vander Woude, K. Aging and problem solving. *Archives of General Psychiatry*, Feb., 1969, 20.

Scheffé, H. A method for judging all possible constrasts in the analysis of variance. *Biometrika*, 1953, 40, 87-104.

Spearman, C. E. *Abilities of Man*. New York, MacMillan, 1927.

Spearman, C. E. *The Nature of "Intelligence" and the Principles of Cognition*. London, MacMillan, 1923.

Tabor, A. Process analysis of Rorschaeh Interpretation. Loyola University, Chicago, Ill., 1959 (unpublished Ph.D. dissertation).

Thurstone, L. L. Primary mental abilities. *Psychometric Monographs*, No. 1 (1938), University of Chicago Press.

Vander Woude, K. Problem solving and language. *Archives of General Psychiatry* (in press), 1970.

APPENDIX A

$31P_1$

Jimmy's grandfather gave Jimmy some big and small balls. They were of two colors, blue and gray. Jimmy misplaced one kind. Which one did he misplace -- the big or the small, the blue or the gray?

	Questions*	Answers
1.	Airplane	No
2.	Big gray balls	No
3.	Black balls	No
4.	Big and small gray balls	No
5.	Wagons	No
6.	Boat	No
7.	Yellow balls	No
8.	Small gray balls	No
9.	Ponies	No
10.	Big blue balls	No

*In all the P problems the "questions" are drawings presented on cards. The description of the corresponding drawing is given next to the number of the card.

$33P_1$

Bob went to a boat show. There were 3 kinds of boats, passanger boats, sail boats and row boats. They were either gray or blue. One of the boats got a prize. Which one was that -- the passanger, sail or row, gray or blue?

	Questions	Answers
1.	One blue and one gray	No
2.	Yellow sail boats	No
3.	Blue sail boats	No
4.	A gray passanger boat	No
5.	A blue row boat	No
6.	A gray sail boat	No
7.	Airplanes	No
8.	A blue passanger boat	No
9.	Blue cars	No
10.	A blue and a gray sail boat.	No

$35P_1$

Mark had saved enough money to buy a new toy, so he went to the toy store. He saw boats, cars and trucks. They were either white or gray or black. He wanted to buy only one. Which one did he buy?

Questions	Answers
1. White, gray, black boats	No
2. A chair	No
3. A white car	No
4. White, gray, black cars	No
5. Horses	No
6. Wagons	No
7. White and black trucks	No
8. A yellow boat	No
9. A white boat and white car	No
10. A gray boat	No

31P$_2$

There are two kinds of trains, passanger and freight. They have either two cars or six cars. I am thinking of one of these. Could you tell me which one I am thinking about?

	Questions	Answers
1.	A passanger and a freight train with 4 cars each	No
2.	A passanger-freight train with 5 cars.	No
3.	A 2-car and a 6-car freight train	No
4.	An airplane	No
5.	A freight train with 2 cars	No
6.	A freight train with 6 cars	No
7.	A passanger train with 2 cars	No
8.	An automobile	No
9.	A passanger and a freight train both with 8 cars	No
10.	A bus	No

$33P_2$

John had two kinds of squirt guns -- big and little. They were yellow squirt guns and black squirt guns. He gave one of these to his little brother. Which one did he give -- the black or the yellow, the big or the little?

	Questions	Answers
1.	A little blue squirt gun	No
2.	A little yellow squirt gun	No
3.	A large and a small yellow squirt gun	No
4.	A large and a small orange squirt gun	No
5.	A little black squirt gun	No
6.	A bow and an arrow	No
7.	A black and a yellow large squirt gun	No
8.	A big black squirt gun	No
9.	A big and a small black squirt gun	No
10.	A big yellow squirt gun	No

$35P_2$

Bill wanted something cool for the afternoon snack. So he went to the ice cream shop. At the shop they had ice cream cones, snow cones and popsickles. They had 3 flavors, strawberry, lemon and chocolate. Which one did Bill buy -- an ice cream cone, a snow cone or a popsickle, and what flavor?

	Questions	Answers
1.	Ice cream cone	No
2.	An orange popsickle	No
3.	A lemon ice cream cone, a strawberry popsickle and a blueberry snow cone	No
4.	A chocolate ice cream cone	No
5.	Popsickles: orange, lemon strawberry and blueberry	No
6.	A strawberry snow cone	No
7.	A strawberry and a lemon snow cone	No
8.	A blueberry popsickle	No
9.	A blueberry snow cone	No
10.	Lemon, strawberry, blueberry and orange ice cream cones	No

$31P_3$

Don has 2 sets of blocks, big and small. There are some yellow and some blue. Yesterday he was playing with only one kind. Could you find out what kind -- big or small, blue or yellow?

	Questions	Answers
1.	An airplane	No
2.	Big yellow blocks	No
3.	Small orange blocks	No
4.	Large and small yellow blocks	No
5.	Wagons	No
6.	A gray boat	No
7.	Little blue blocks	No
8.	Little yellow blocks	No
9.	Horses	No
10.	Big brown blocks	No

$33P_3$

John had 3 kinds of boats, canoes, sail boats and row boats. They were brown and black. He lost one of these boats. Which one did he lose -- a brown boat or a black boat, a canoe or a row boat, or a sail boat?

Questions	Answers
1. A brown and a black canoe	No
2. Yellow sail boats	No
3. Brown sail boats	No
4. A black canoe	No
5. A brown row boat	No
6. A black sail boat	No
7. Airplanes	No
8. A brown canoe	No
9. Blue cars	No
10. A brown and a black sail boat	No

$35P_3$

Tom has 3 kinds of automobiles -- buses, trucks and cars. They are white, gray and black. He uses one of these for his personal use. Which one does he use -- the car, bus or the truck and of what color, gray, black or white?

Questions	Answers
1. Black and gray buses	No
2. A chair	No
3. A white car	No
4. White, gray and black cars	No
5. A horse	No
6. Wagons	No
7. White, gray and black truck	No
8. A yellow bus	No
9. A gray bus and a white car	No
10. A gray truck	No

31VA$_1$

Instructions and Corresponding Questions and Answers

John has 20 horses. There are black race horses and white race horses. There are black farm horses and white farm horses. I want you to figure out how many black farm horses there are?

Questions	Answers
1. How many horses does John ride?	1. 10.
2. How many white horses does John have?	2. 7.
3. How many brown horses does John have?	3. 0.
4. How many white racing horses does John have?	4. 5.
5. How many black racing horses does John have?	5. 5.
6. How many brown racing horses does John have?	6. 0.
7. How many white farm horses does John have?	7. 2.
8. How many brown farm horses does John have?	8. 0.
9. How many horses did John sell?	9. 0.
10. How many ponies does John have?	10. 0.

BVA_1

Instructions and Corresponding Questions and Answers

Daddy bought Jimmy 50 trucks. Some were red, some blue and some green. Some were fire trucks and some were army trucks. How many green army trucks did Daddy buy Jimmy.

Questions	Answers
1. How many cars does Daddy have for himself?	1. 1.
2. How many cars did Daddy buy Jimmy?	2. 0.
3. How many red trucks were there?	3. 15.
4. How many blue trucks were there?	4. 17.
5. How many red fire trucks were there?	5. 3.
6. How many red army trucks were there?	6. 12.
7. How many blue fire trucks were there?	7. 7.
8. How many blue army trucks were there?	8. 10.
9. How many green fire trucks were there?	9. 7.
10. How many brown trucks did Daddy buy Jimmy?	10. 0.

$35VA_1$

Instructions and Corresponding Questions and Answers

Joe and his two friends Peter and Mark went to the store to buy some marbles. Each one of them bought some green ones, some red ones, and some blue ones. Altogether they bought 45 marbles. How many blue marbles did Mark buy?

Questions	Answers
1. How many green marbles did the three of them buy?	1. 15.
2. How many red marbles and green marbles did Peter buy?	2. 10.
3. Did they use the marbles right away?	3. Yes.
4. How many green marbles did Mark buy?	4. 5.
5. How many red marbles did Peter buy?	5. 5.
6. Did Peter buy more marbles than Joe?	6. No.
7. Are the red marbles larger than the green ones?	7. No.
8. How many blue marbles did Jow and Peter buy?	8. 10.
9. Did they buy anything else besides marbles?	9. No.
10. How many red marbles did the three of them buy?	10. 15.

31VA$_2$

Instructions and Corresponding Questions and Answers

Last Sunday the second graders had a party for all the boys and girls in the class. There were 10 boys and girls at the party and some wanted orange pop and some root beer. How many boys drank root beer?

Questions	Answers
1. Did all the boys and girls go to the party?	1. Yes.
2. How many girls drank root beer?	2. 2.
3. How many bottles of pop did they have at the party?	3. 24.
4. Was the teacher at the party?	4. Yes.
5. Was the party in the morning or in the afternoon?	5. Afternoon.
6. How many girls were at the party?	6. 6.
7. Did they play games at the party?	7. Yes.
8. How many girls drank orange pop?	8. 4.
9. Were there more boys than girls at the party?	9. No.
10. How many boys drank orange pop?	10. 1.

33VA$_2$

Instructions and Corresponding Questions and Answers

David and his two friends Bob and Dan decided to go fishing. Each of the boys caught some blue gils and some perch. At the end of the day the boys had caught a total of 15 fish. How many blue gils did David catch?

Questions	Answers
1. How many perch did Bob catch?	1. 2.
2. Who caught the most fish?	2. Each boy caught the same number of fish.
3. How many fish did Bob catch?	3. 5.
4. How many bull heads did they catch?	4. 0.
5. What kind of fishing poles did they use?	5. Cane poles.
6. How many blue gils did Dan catch?	6. 3.
7. What kind of bait did they use?	7. Minnows.
8. How many fish did Dan catch?	8. 5.
9. How much time did the boys spend fishing?	9. 4 hours.
10. How many perch did David catch?	10. 2.

$35VA_2$

Instructions and Corresponding Questions and Answers

Joe and Jack and Jim each play on the same baseball team. At the end of the season the three boys found that together they had a total of 90 home runs, walks and strikeouts. How many home runs did Jack have?

Questions	Answers
1. How many walks did Jim have?	1. 10.
2. How many home runs, walks and strikeouts did Jim have?	2. 30.
3. How many strikeouts did Jack have?	3. 10.
4. Did each have the same number of walks?	4. No.
5. How many walks and strikeouts did Jack have?	5. 20.
6. Who had the most strikeouts?	6. Joe.
7. How many home runs did Joe have?	7. 10.
8. Who had the most walks?	8. Joe and Jack each had 10.
9. How many strikeouts, walks and home runs did Joe have.	9. 30.
10. Who had the most homeruns?	10. Jim.

31VA$_3$

Instructions and Corresponding Questions and Answers

At the dog show this year, there were 20 dogs entered. There were some large black dogs and some large brown dogs; there were also some small black dogs and small brown dogs. See if you can figure out how many large black dogs there were in the show.

Questions	Answers
1. What day was the dog show?	1. Saturday.
2. How many small dogs were there in the show?	2. 7.
3. How many white dogs were there?	3. None.
4. How many small black dogs were there?	4. 5.
5. How many large brown dogs were there?	5. 5.
6. Were there any poodles?	6. Yes.
7. How many small brown dogs were there?	7. 2.
8. How many small white dogs were in the show?	8. None.
9. How many prizes were given?	9. 3.
10. How many large white dogs were there?	10. None.

33VA$_3$

Instructions and Corresponding Questions and Answers

John and Billy went to the carnival and bought 40 tickets for rides. Some were for the Ferris Wheel, some for the Roller Coaster, and some for the Merry-go-Round. They each bought some tickets for all three rides. How many tickets did John get for the Merry-go-Round?

Questions	Answers
1. Did John's dad go with them to the carnival?	1. Yes.
2. How much popcorn did they buy at the carnival?	2. They each had one box.
3. How many Ferris Wheel tickets did the two of them buy?	3. 14.
4. How many Roller Coaster tickets did the boys buy?	4. 16.
5. How many Ferris Wheel tickets did Billy buy?	5. 8.
6. How many Ferris Wheel tickets did John buy?	6. 6.
7. How many Roller Coaster tickets did Billy buy?	7. 6.
8. How many Roller Coaster tickets did John buy?	8. 10.
9. How many Merry-go-Round tickets did Billy buy?	9. 4.
10. How many Bingo games did John's dad play?	10. 3.

35VA$_3$

Instructions and Corresponding Questions and Answers

Jimmy and his friends decided to count the commercials they saw on TV for one week. The shows they watched were "Flipper", "Lassie", and "Batman". Each show had some commercials for toothpaste, some for cereal, and some for candy. Altogether they counted 45 commercials. How many candy commercials did they see on "Batman".

Questions	Answers
1. How many cereal commercials were there altogether?	1. 20.
2. How many cereal and toothpaste commercials did they see on "Flipper"?	2. 10.
3. What day did they start counting the commercials?	3. Sunday.
4. How many cereal commercials were on "Batman"?	4. 10.
5. How many toothpaste commercials were on "Flipper"?	5. 2.
6. Were there more commercials on "Batman" than on "Lassie"?	6. Yes.
7. Were the toothpaste commercials better than the cereal commercials?	7. No.
8. How many candy commercials were on "Flipper" and "Lassie" together?	8. 11.
9. Which was their favorite show?	9. "Lassie".
10. How many toothpaste commericals were there in all?	10. 10.

$31VB_1$ ($31VB_2$)

Instructions and Corresponding Questions and Answers

We have 50 objects called C. There are two kinds of C's, one kind is called B, the other kind is called G. Any B can be either a R or a T, and any G can be either a R or a T. No B can be a G and no R can be a T. No B can be a G and no R can be a T. Will you find out how many of the G objects are also called T?

Questions	Answers
1. How many K's are there?	1. 11.
2. How many H objects are also called G?	2. 15.
3. How many T objects are also called B?	3. 10.
4. How many N objects are there?	4. 10.
5. How much is K times C?	5. 550.
6. Are there more G than B objects?	6. No.
7. How many R objects are there?	7. 35.
8. Are there more R objects than T objects?	8. Yes.
9. Are there any objects called M?	9. No.
10. How many R objects are also called B?	10. 20.

$33VB_1$ ($33VB_2$)

Instructions and Corresponding Questions and Answers

There are 40 objects called T. There are two kinds of T objects, one kind is called E and the other kind is called F. Each F object is either an A, a S or a P. Also each E object is either an A, a S or a P. No E object is called F and no F objects are called E. Also no A object is called S or P, no S object is called A or P and no P object is called S or A. How many A objects are also called E.

Questions	Answers
1. Are any objects called K?	1. No.
2. How many E objects are also called P?	2. 15.
3. How many objects are called S?	3. 10.
4. How many E objects are also called S?	4. 4.
5. How many objects are called P?	5. 20.
6. How many F objects are also called A?	6. 8.
7. How many F objects are also called P?	7. 5.
8. How many E objects are also called W?	8. 0.
9. How many F objects are also called S?	9. 6.
10. Are there more P objects than S objects?	10. Yes.

$35VB_1$ ($35VB_2$)

Instructions and Corresponding Questions and Answers

We have three kinds of T objects. One kind is called M, another kind is called N, and another kind is called P. Further, each M, N, or P can also be called either a Q, a R, or an S. Altogether there are fifty objects. How many of the N objects are also called S?

	Questions		Answers
1.	How many Q objects and R objects are called P?	1.	15.
2.	How many M objects and P objects are also called S?	2.	5.
3.	Are there more Q objects than S objects?	3.	Yes.
4.	How many N objects are called Q?	4.	5.
5.	How many objects are called Q?	5.	25.
6.	How many M objects are called A?	6.	0.
7.	How many objects are called R?	7.	15.
8.	Are there more P objects than R objects?	8.	Yes.
9.	How many objects are called K?	9.	0.
10.	How many P objects are also called R?	10.	5.

$31VB_3$

Instructions and Corresponding Questions and Answers

We have 40 objects called S. There are two kinds of S's. One kind is called P; the other kind is called Q. Any P can be an A or a B, and any Q can be an A or a B. No P can be a Q, and no A can be a B. Will you find out how many of the Q objects are also called B?

Questions	Answers
1. How many K's are there?	1. 7.
2. How many A objects are also called Q?	2. 8.
3. How many B objects are also called P?	3. 14.
4. How many N objects are there?	4. 5.
5. How much is K times S?	5. 280.
6. Are there more Q than P objects?	6. No.
7. How many A objects are there?	7. 14.
8. Are there more A objects than B objects?	8. No.
9. Are there any objects called M?	9. No.
10. How many A objects are also called P?	10. 6.

33VB$_3$

Instructions and Corresponding Questions and Answers

There are 30 objects called N. There are two kinds of N objects, one kind is called H and the other kind is called L. Each L object is either an X, a Y or a Z. Also each H object is either an X, a Y or a Z. No H object is called L and no L objects are called H. Also no X object is called Y or Z, no Y object is called X or Z, and no Z object is called Y or X. How many X objects are also called H?

Questions		Answers
1.	Are any objects called K?	1. No.
2.	How many H objects are also called Z?	2. 3.
3.	How many objects are called Y?	3. 13.
4.	How many H objects are also called Y?	4. 7.
5.	How many objects are called Z?	5. 5.
6.	How many L objects are also called X?	6. 8.
7.	How many L objects are also called Z?	7. 2.
8.	How many H objects are also called W?	8. 0.
9.	How many L objects are also called Y?	9. 6.
10.	Are there more Z objects than Y objects?	10. No.

m

$35VB_3$

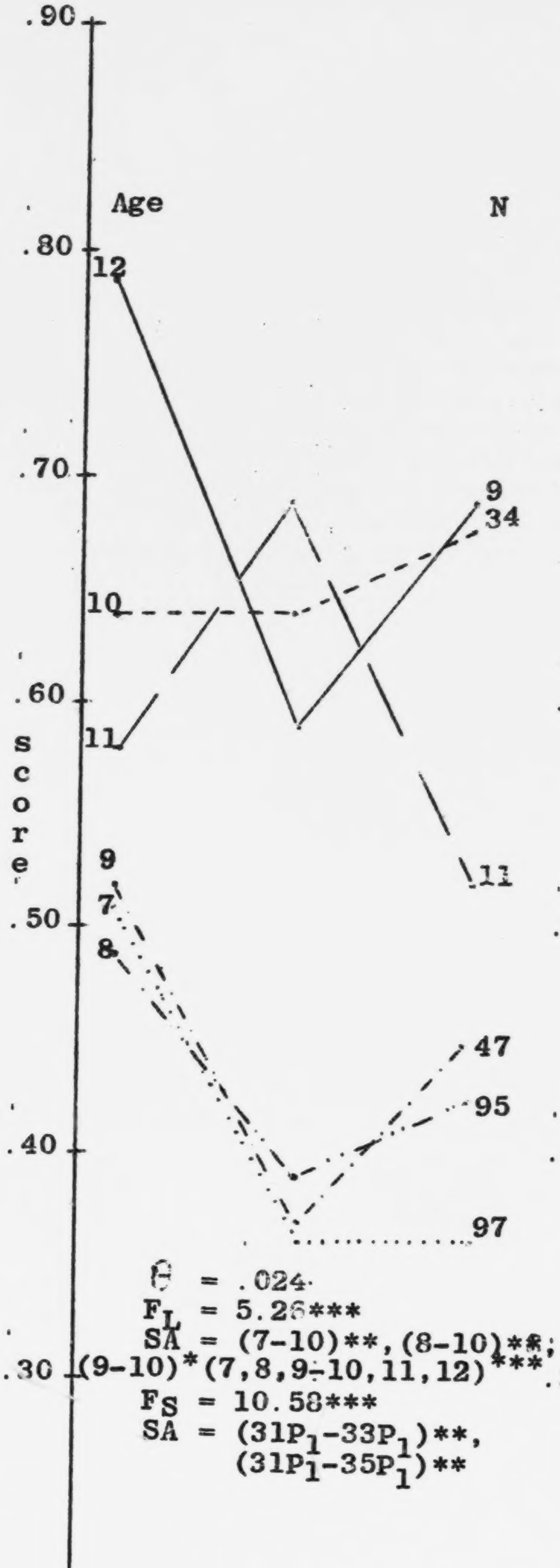
Instructions and Corresponding Questions and Answers

We have three kinds of G objects. One kind is called A, another kind is called B, and another kind is called C. Further, each A, B, or C can also be called either an M, an N, or a Q. Altogether there are sixty objects. How many of the B objects are also called Q?

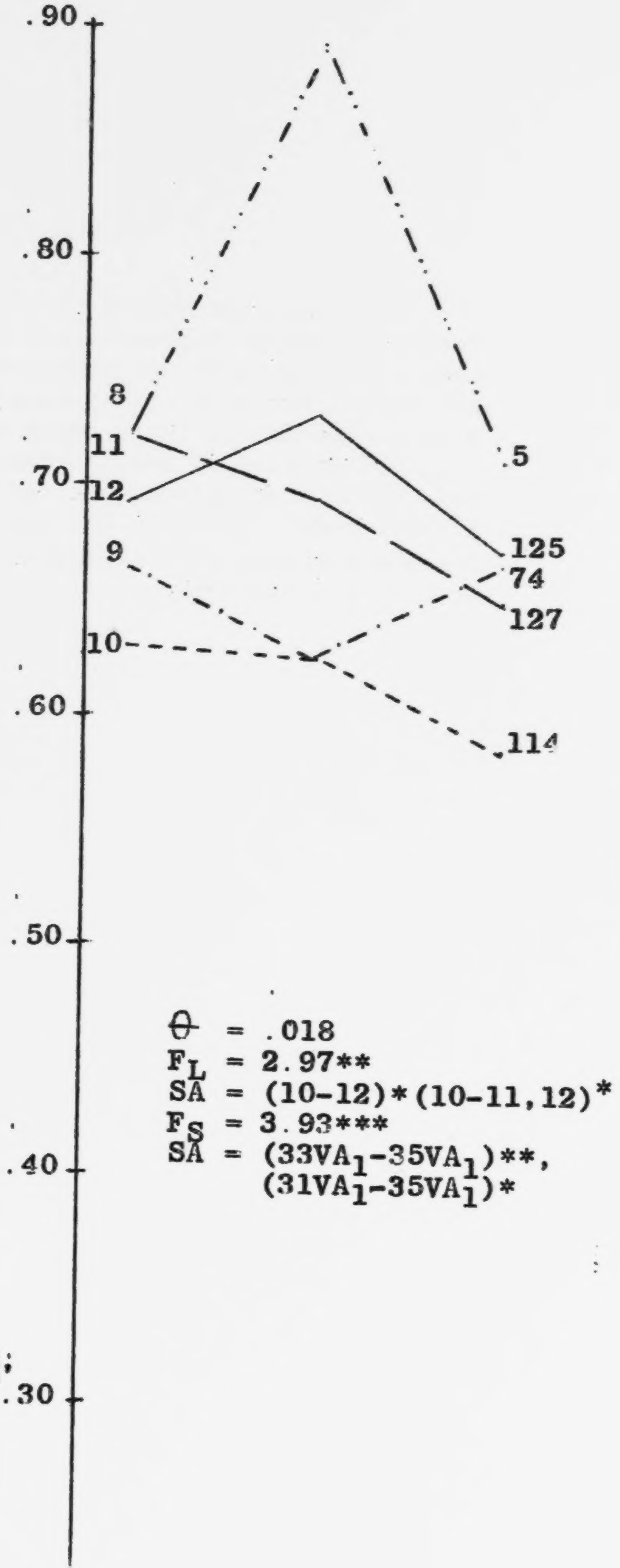
	Questions		Answers
1.	How many M objects and N objects are called C?	1.	14.
2.	How many A objects and C objects are also called Q?	2.	7.
3.	Are there more M objects than Q objects?	3.	Yes.
4.	How many B objects are called M?	4.	7.
5.	How many objects are called M?	5.	32.
6.	How many A objects are called F?	6.	0.
7.	How many objects are called N?	7.	14.
8.	Are there more C objects than N objects?	8.	Yes.
9.	How many objects are called H?	9.	0.
10.	How many C objects are also called N?	10.	2.

APPENDIX B

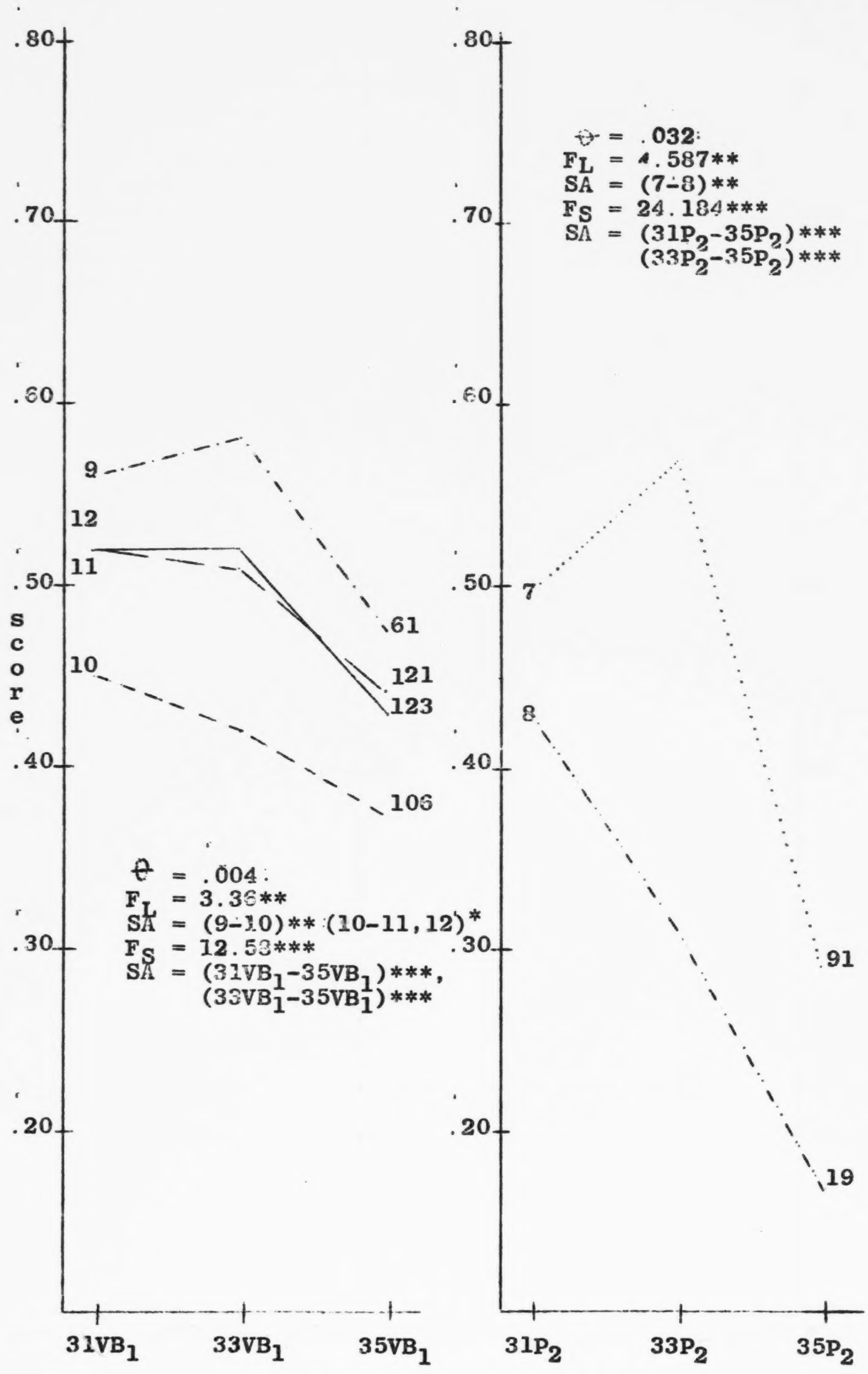
The graphs presented here illustrate the multivariate results given in Tables 21 and 22. Symbols to the left of each curve identify a specific Age or I.Q. group; those to the right, the size of the sample. The parallelism statistic, θ indicates in all these cases that parallelism could not be rejected at the .05 level. Significance levels for F-ratios and Scheffé Comparisons are denoted "*" for the .10 level of significance, "**" for .05, and "***" for .01. F_L is the F-ratio for ages or I.Q. levels, and F_S for the scales (problems or administrations.)

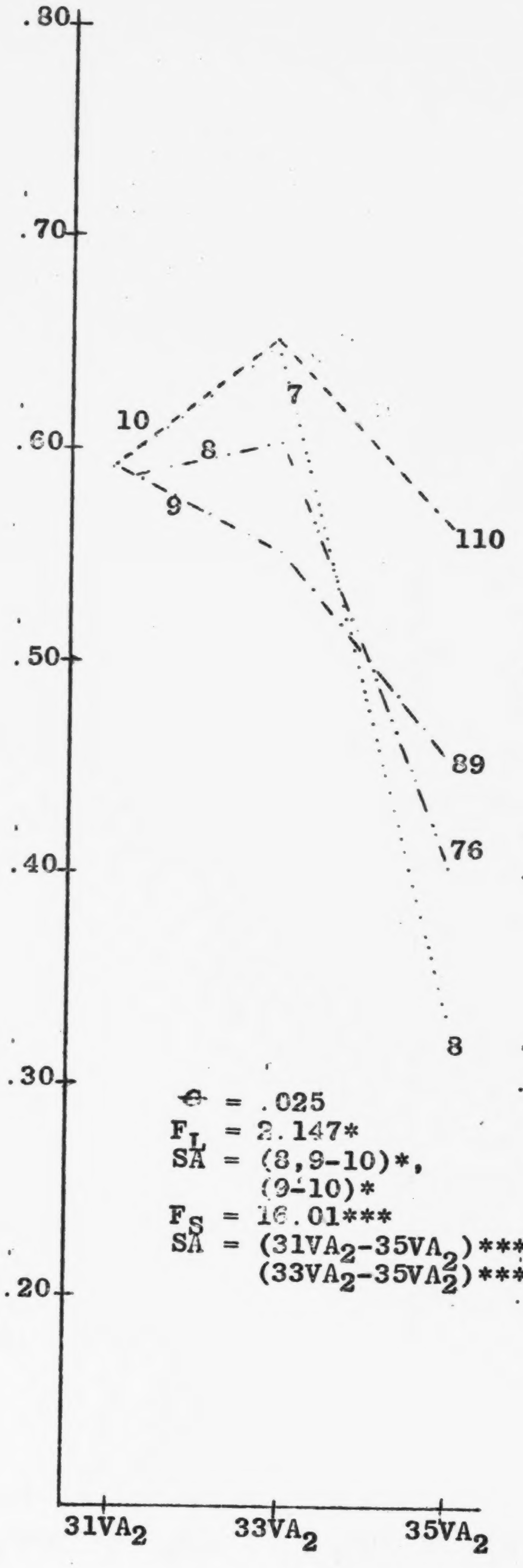


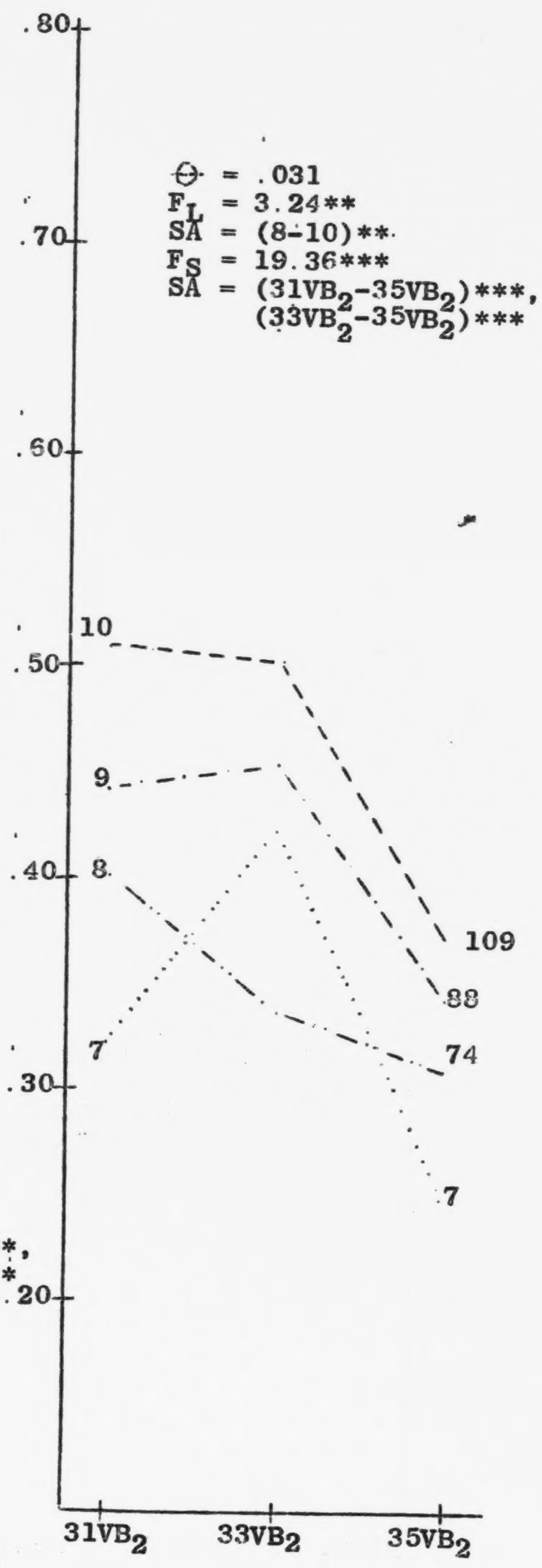
θ = .024
F_L = 5.26***
SA = (7-10)**, (8-10)**;
(9-10)*(7,8,9-10,11,12)***
F_S = 10.58***
SA = ($31P_1$-$33P_1$)**,
($31P_1$-$35P_1$)**

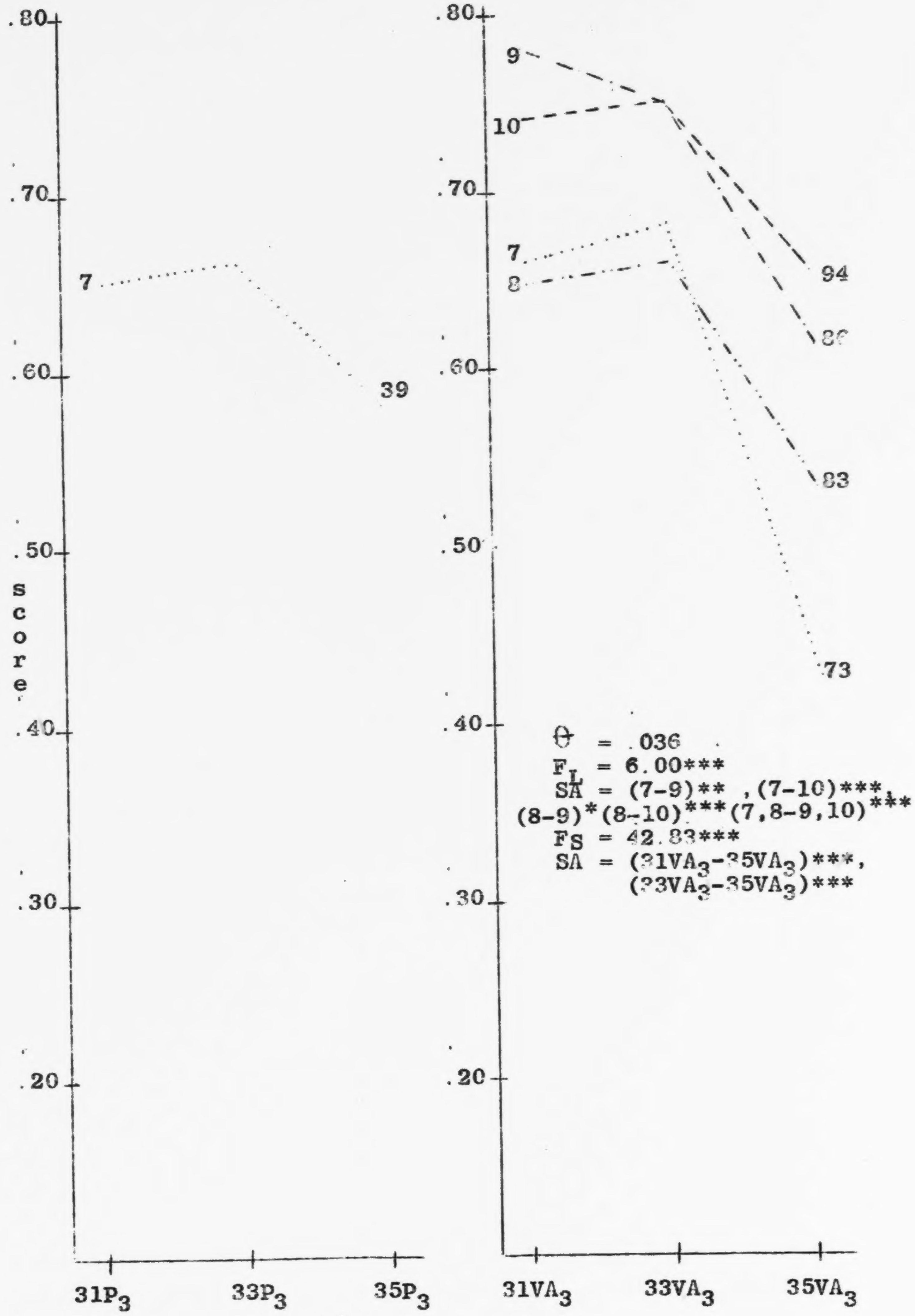


θ = .018
F_L = 2.97**
SA = (10-12)*(10-11,12)*
F_S = 3.93***
SA = ($33VA_1$-$35VA_1$)**,
($31VA_1$-$35VA_1$)*

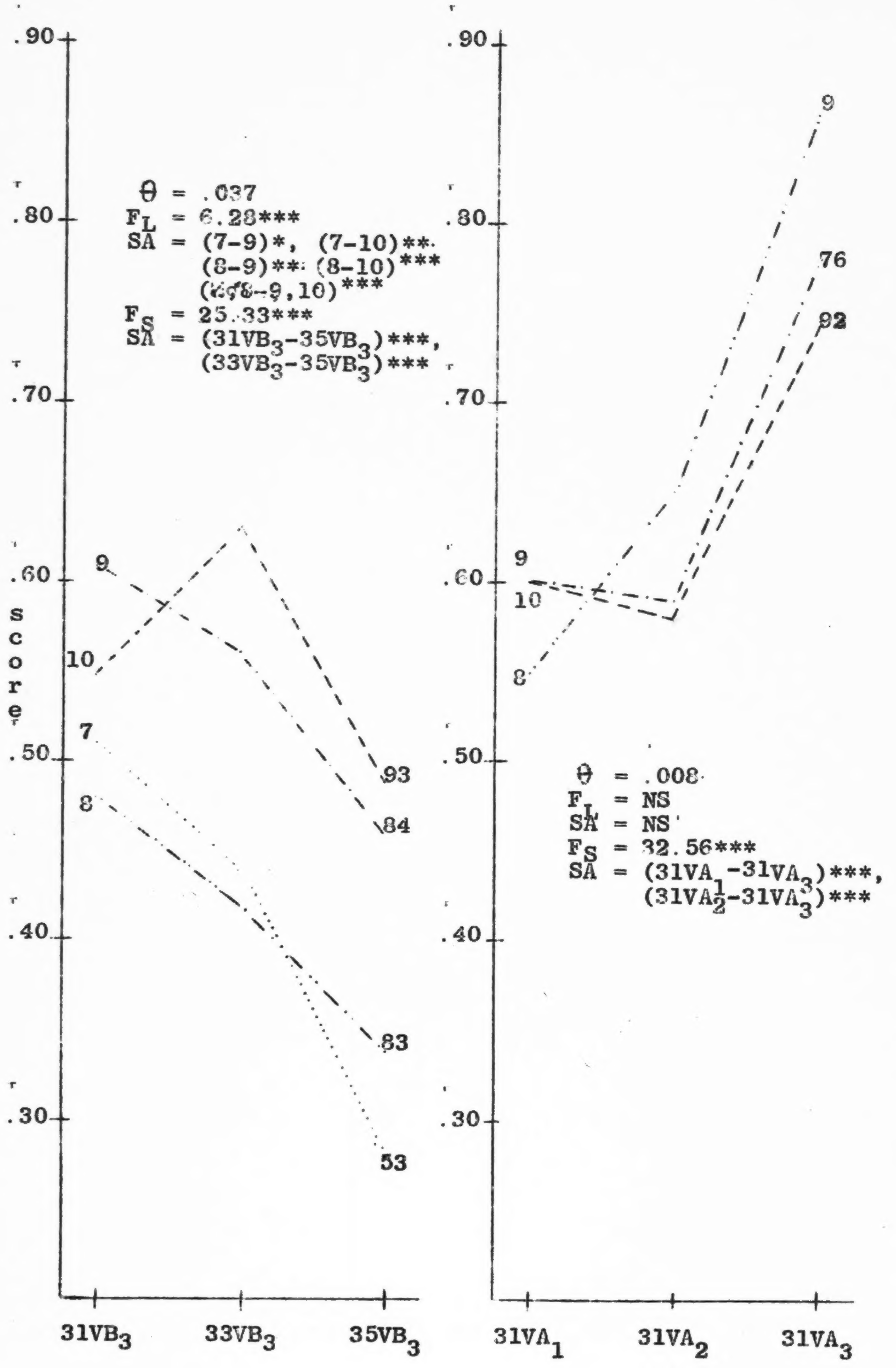
.80
.70
.60
.50
.40
.30
.20
score
9
12
11
10
61
121
123
106
θ = .004
F_L = 3.36**
SA = (9-10)** (10-11,12)*
F_S = 12.53***
SA = (31VB_1-35VB_1)***,
(33VB_1-35VB_1)***
31VB_1
33VB_1
35VB_1
.80
.70
.60
.50
.40
.30
.20
θ = .032
F_L = 4.587**
SA = (7-8)**
F_S = 24.184***
SA = (31P_2-35P_2)***
(33P_2-35P_2)***
7
8
91
19
31P_2
33P_2
35P_2

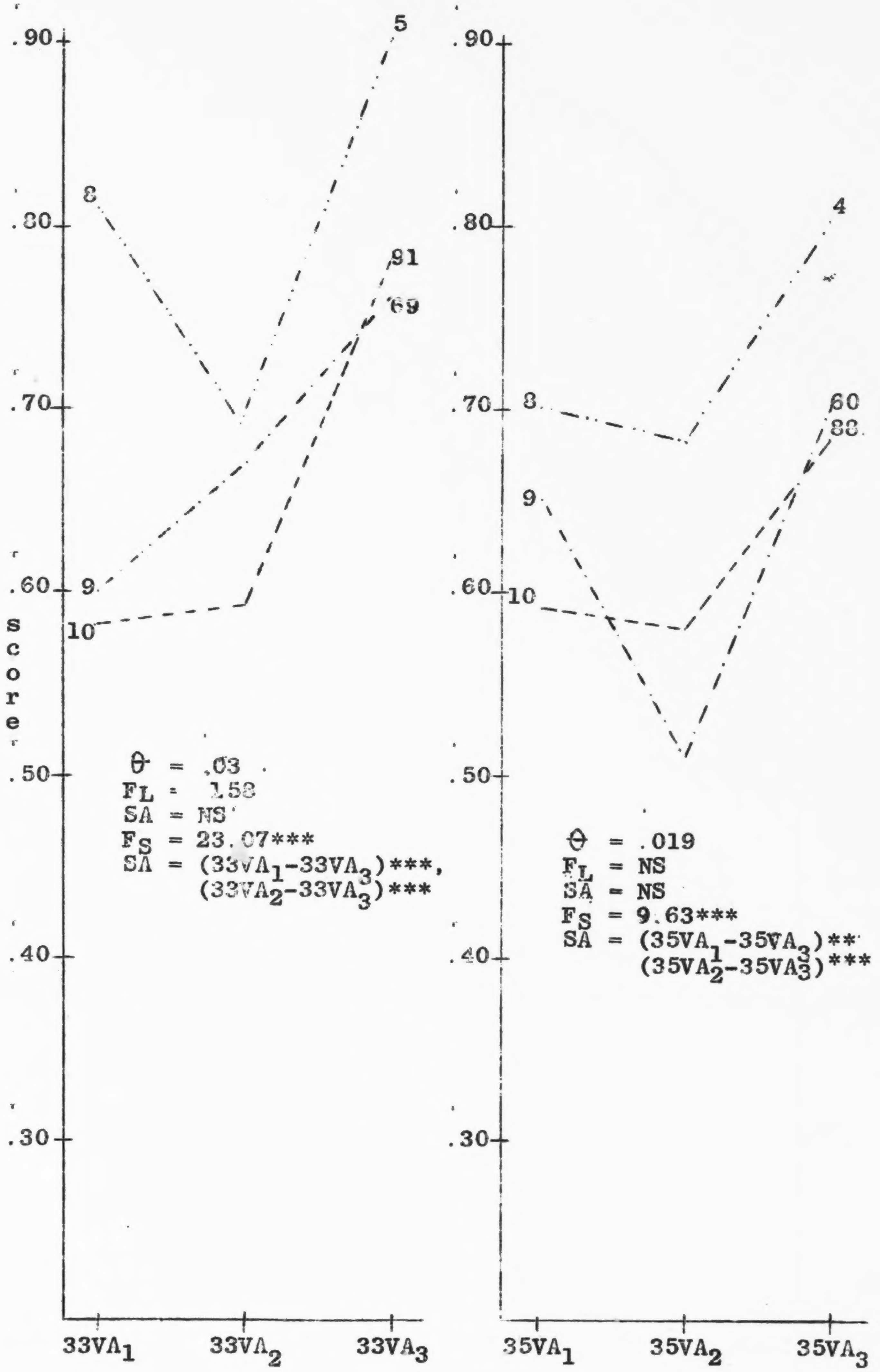

.80
.70
.60
.50
.40
.30
.20
31VA2
33VA2
35VA2
10
8
9
7
110
89
76
8
θ = .025
FL = 2.147*
SA = (8,9-10)*,
(9-10)*
FS = 16.01***
SA = (31VA2-35VA2)***,
(33VA2-35VA2)***


.80
.70
.60
.50
.40
.30
.20
31VB2
33VB2
35VB2
θ = .031
FL = 3.24**
SA = (8-10)**
FS = 19.36***
SA = (31VB2-35VB2)***,
(33VB2-35VB2)***
10
9
8
7
109
88
74
7



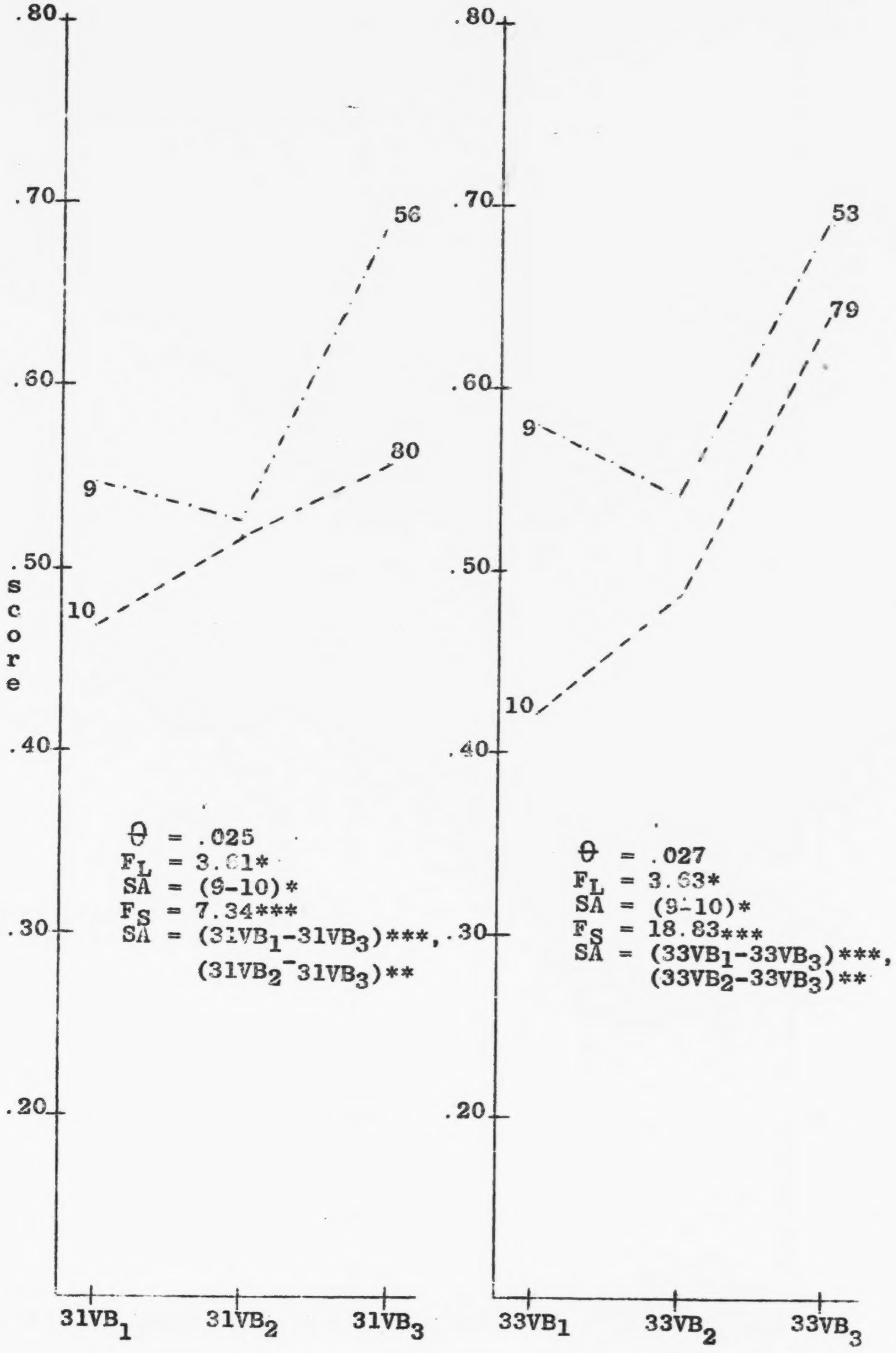
θ = .036
F_L = 6.00***
SA = (7-9)** ,(7-10)***,
(8-9)*(8-10)***(7,8-9,10)***
F_S = 42.83***
SA = ($31VA_3$-$35VA_3$)***,
($33VA_3$-$35VA_3$)***

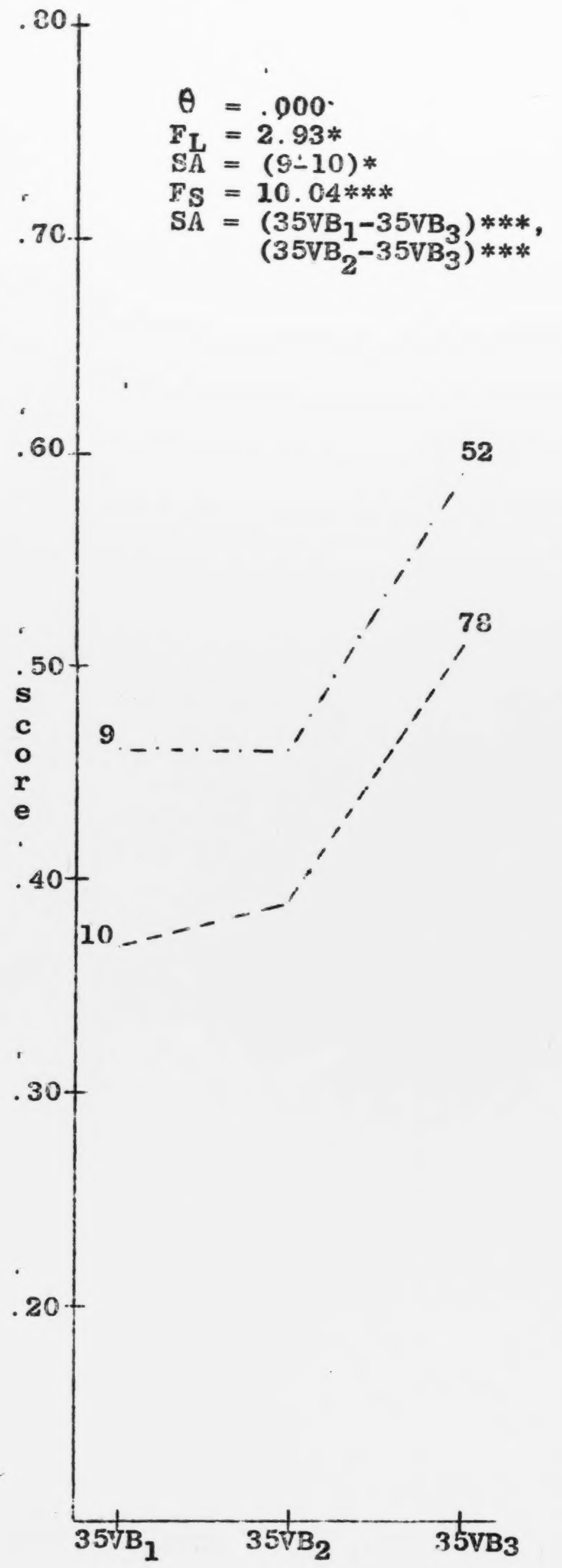

.90
.80
.70
.60
.50
.40
.30
score
θ = .037
F_L = 6.28***
SA = (7-9)*, (7-10)**
(8-9)**: (8-10)***
(7,8-9,10)***
F_S = 25.33***
SA = (31VB3-35VB3)***,
(33VB3-35VB3)***
9
10
7
8
93
84
83
53
31VB3
33VB3
35VB3
.90
.80
.70
.60
.50
.40
.30
9
10
8
9
76
92
θ = .008
F_L = NS
SA = NS
F_S = 32.56***
SA = (31VA1-31VA3)***,
(31VA2-31VA3)***
31VA1
31VA2
31VA3

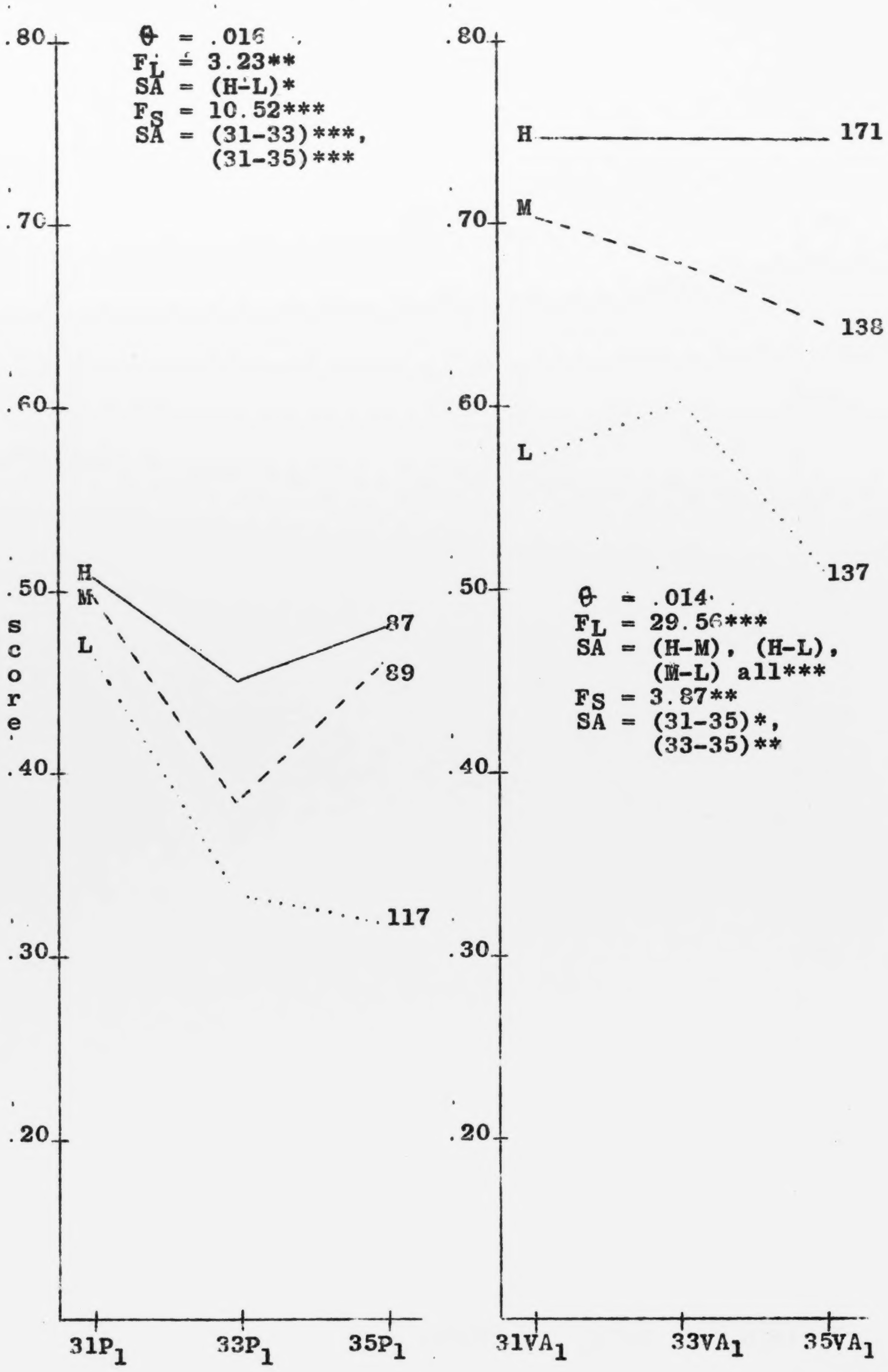


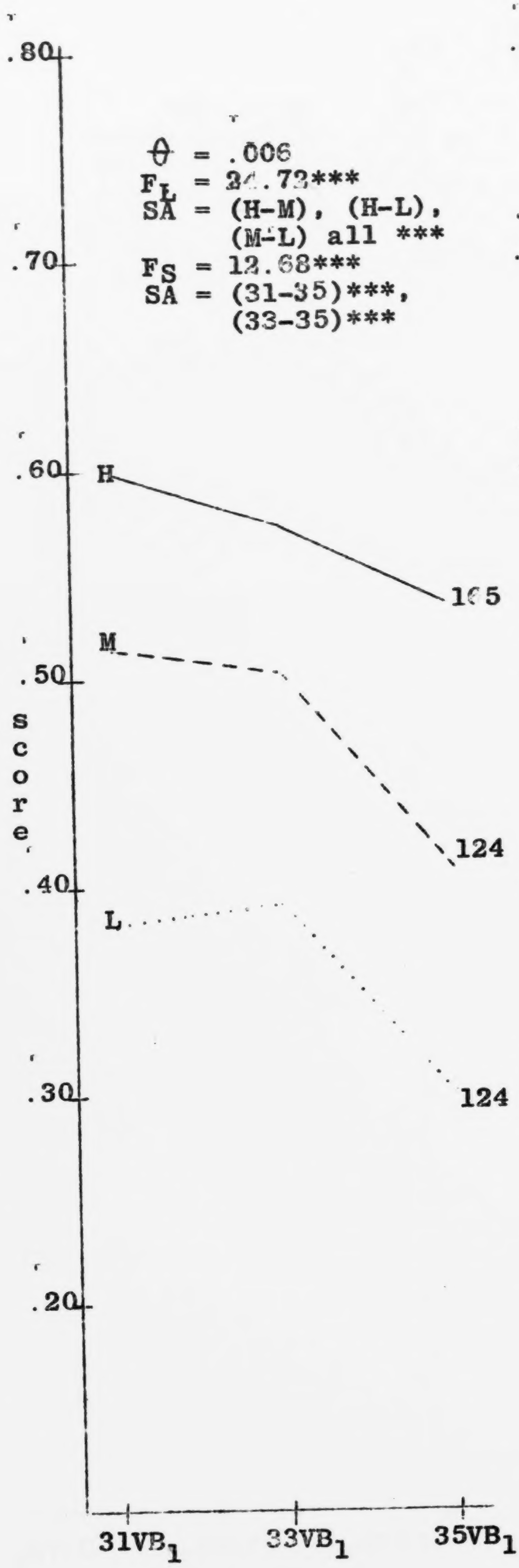
θ = .03
F_L = 158
SA = NS
F_S = 23.07***
SA = $(33VA_1-33VA_3)$***,
$(33VA_2-33VA_3)$***

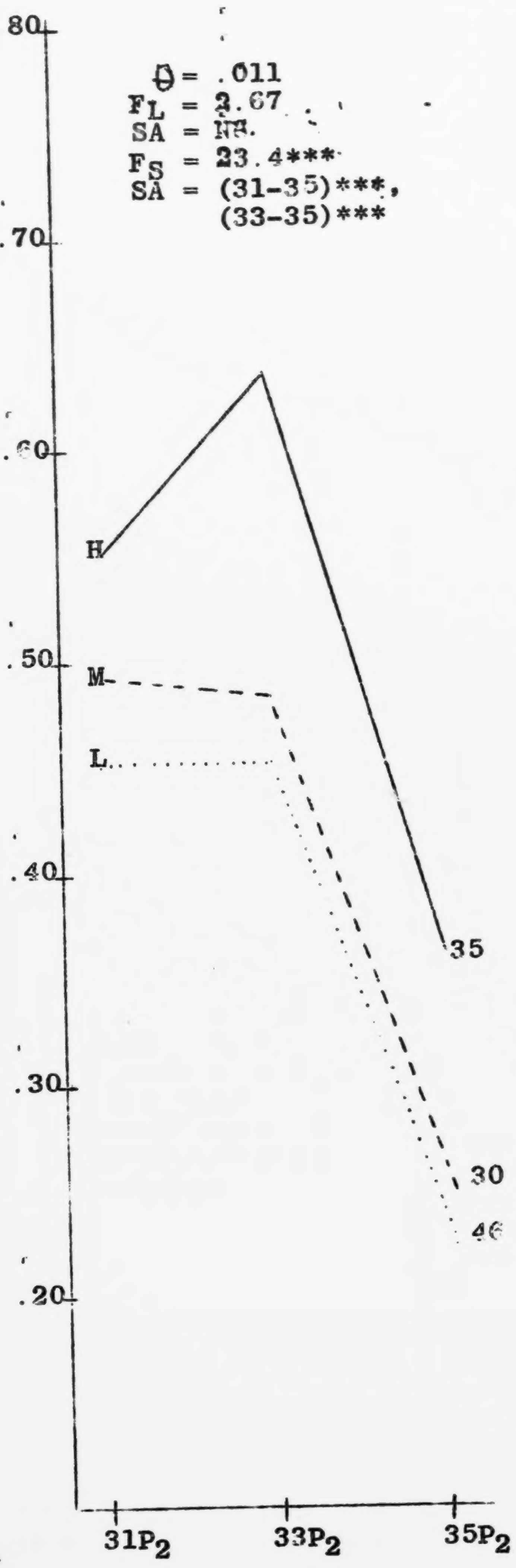
θ = .019
F_L = NS
SA = NS
F_S = 9.63***
SA = $(35VA_1-35VA_3)$**
$(35VA_2-35VA_3)$***

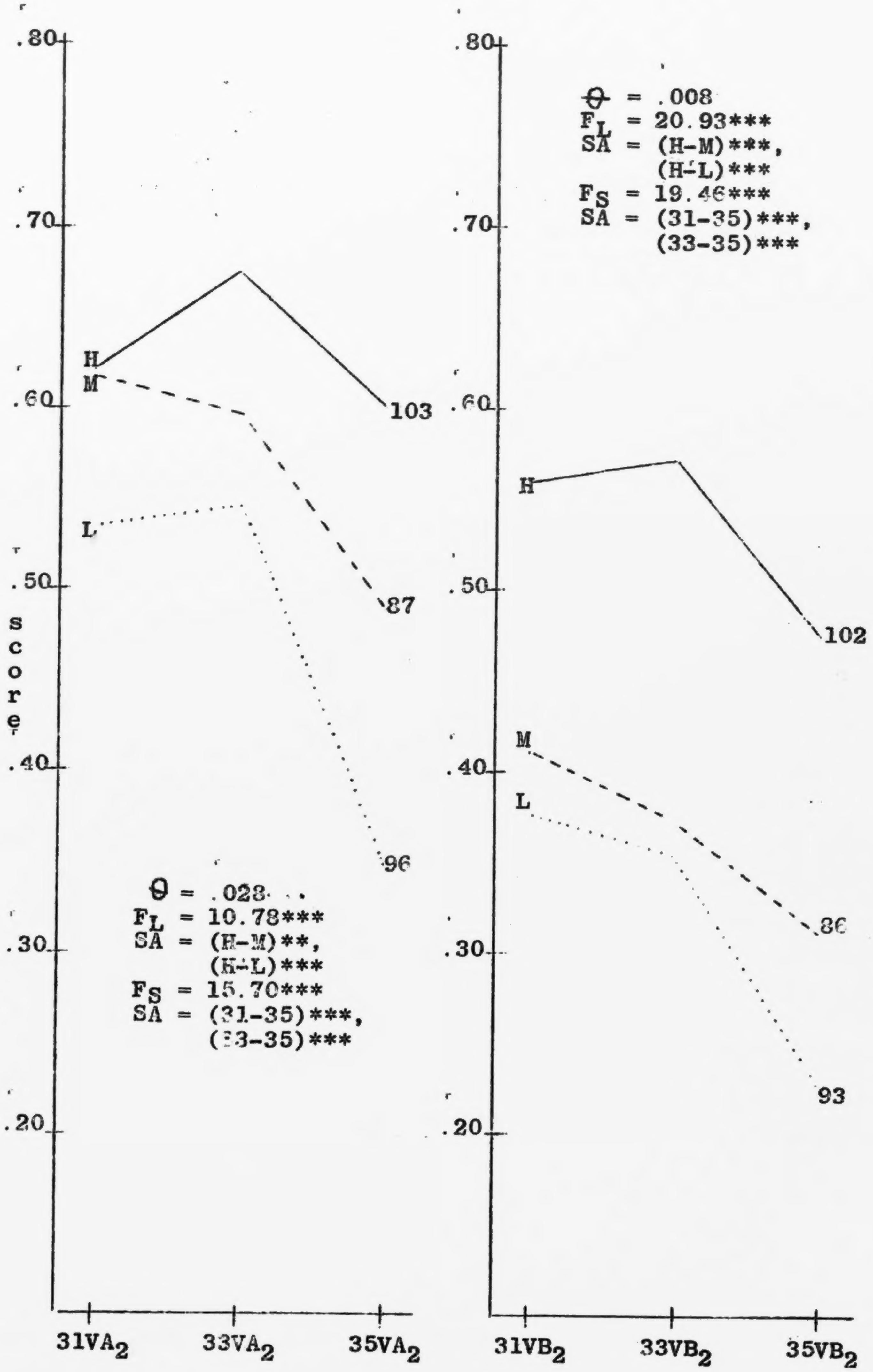

.80
.70
.60
.50
.40
.30
.20
score
9
10
56
80
31VB1
31VB2
31VB3
θ = .025
FL = 3.61*
SA = (9-10)*
FS = 7.34***
SA = (31VB1-31VB3)***,
(31VB2-31VB3)**
.80
.70
.60
.50
.40
.30
.20
9
10
53
79
33VB1
33VB2
33VB3
θ = .027
FL = 3.63*
SA = (9-10)*
FS = 18.83***
SA = (33VB1-33VB3)***,
(33VB2-33VB3)**

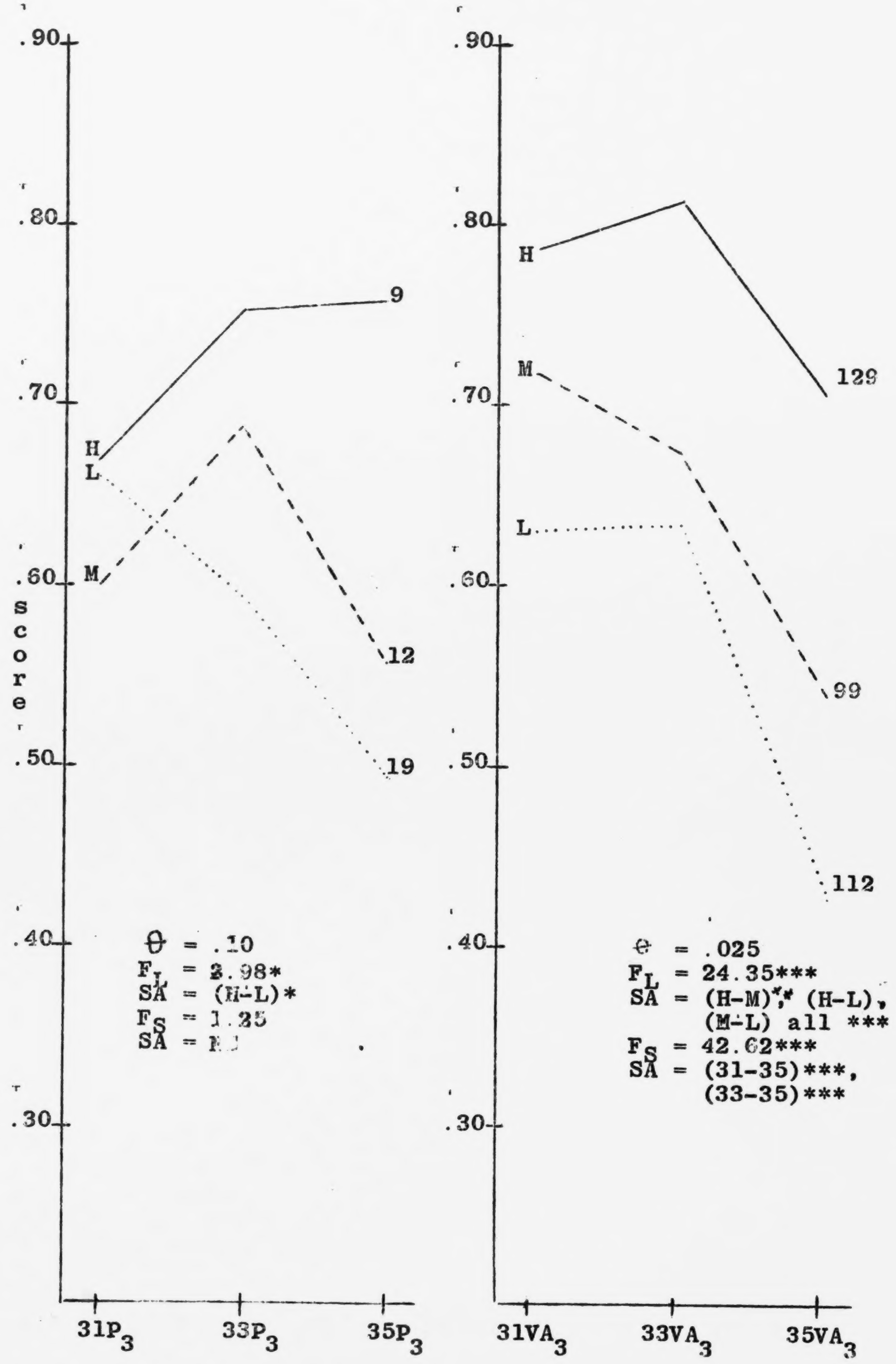

θ = .000
F_L = 2.93*
SA = (9-10)*
F_S = 10.04***
SA = (35VB1-35VB3)***,
(35VB2-35VB3)***
.80
.70
.60
.50
.40
.30
.20
score
9
10
52
78
35VB1
35VB2
35VB3

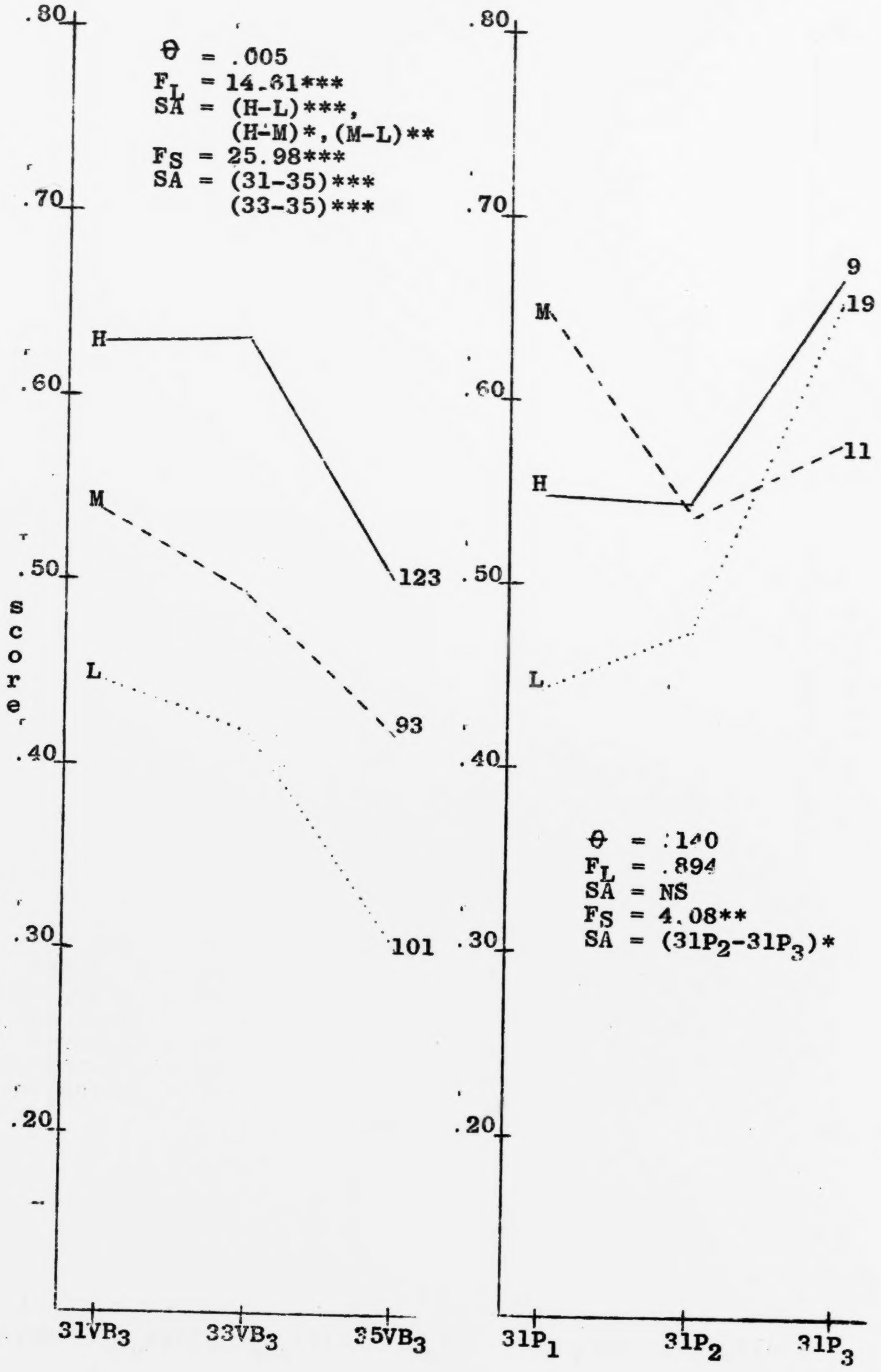

θ = .016
F_L = 3.23**
SA = (H-L)*
F_S = 10.52***
SA = (31-33)***,
(31-35)***
.80
.70
.60
.50
.40
.30
.20
score
H
M
L
87
89
117
31P1
33P1
35P1
.80
.70
.60
.50
.40
.30
.20
H
M
L
171
138
137
θ = .014
F_L = 29.56***
SA = (H-M), (H-L),
(M-L) all***
F_S = 3.87**
SA = (31-35)*,
(33-35)**
31VA1
33VA1
35VA1

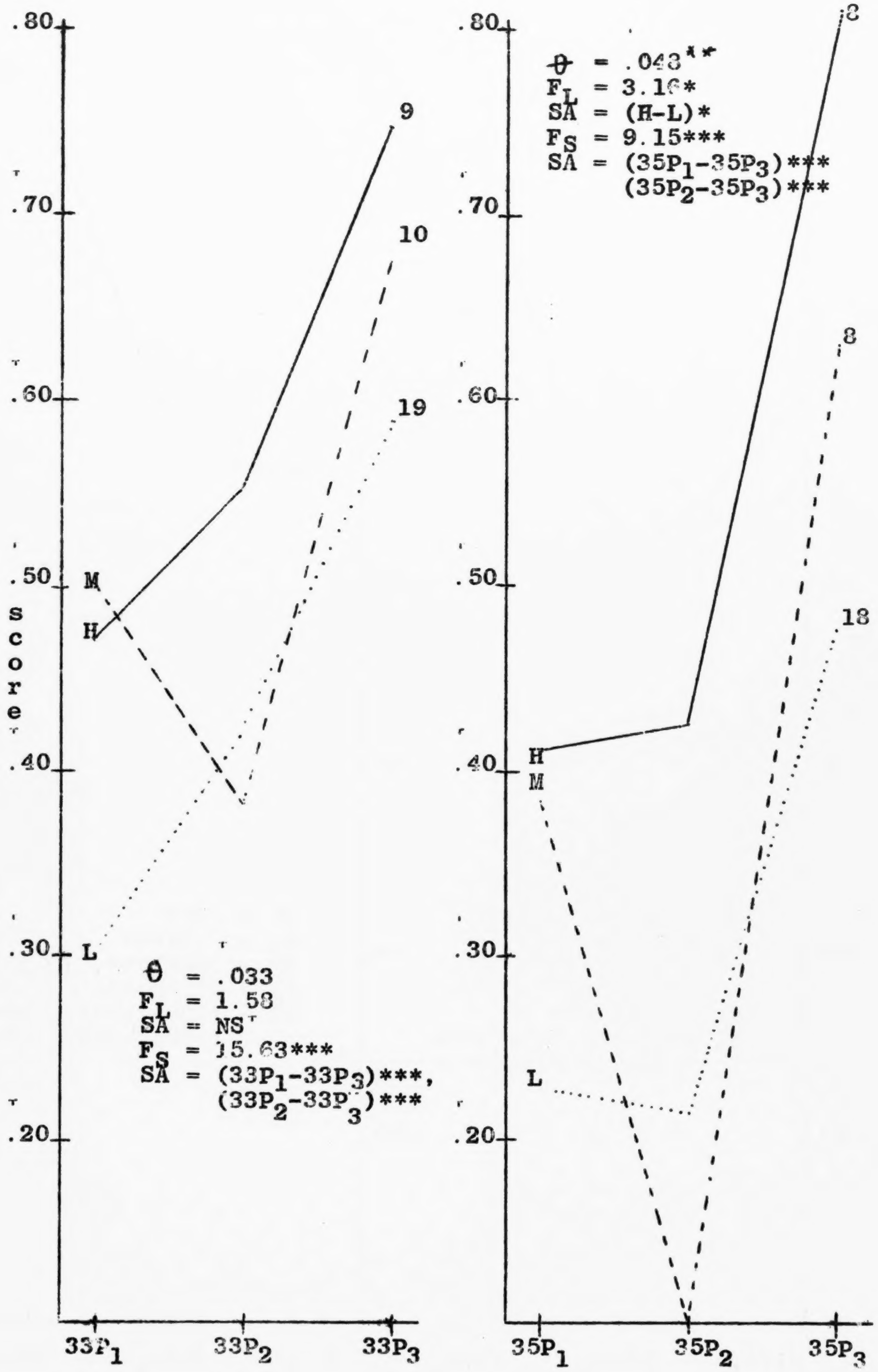

.80
.70
.60
.50
.40
.30
.20
score
θ = .006
F_L = 24.72***
SA = (H-M), (H-L), (M-L) all ***
F_S = 12.68***
SA = (31-35)***, (33-35)***
H
M
L
165
124
124
31VB1
33VB1
35VB1

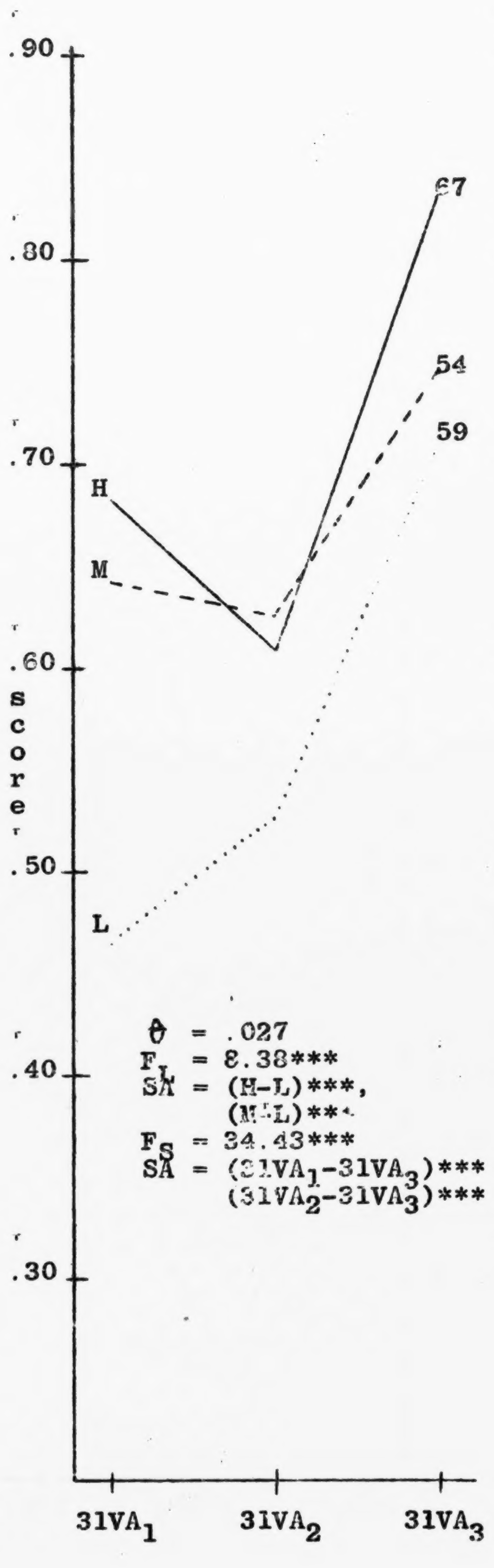

.80
.70
.60
.50
.40
.30
.20
θ = .011
F_L = 3.67
SA = NS.
F_S = 23.4***
SA = (31-35)***, (33-35)***
H
M
L
35
30
46
31P2
33P2
35P2

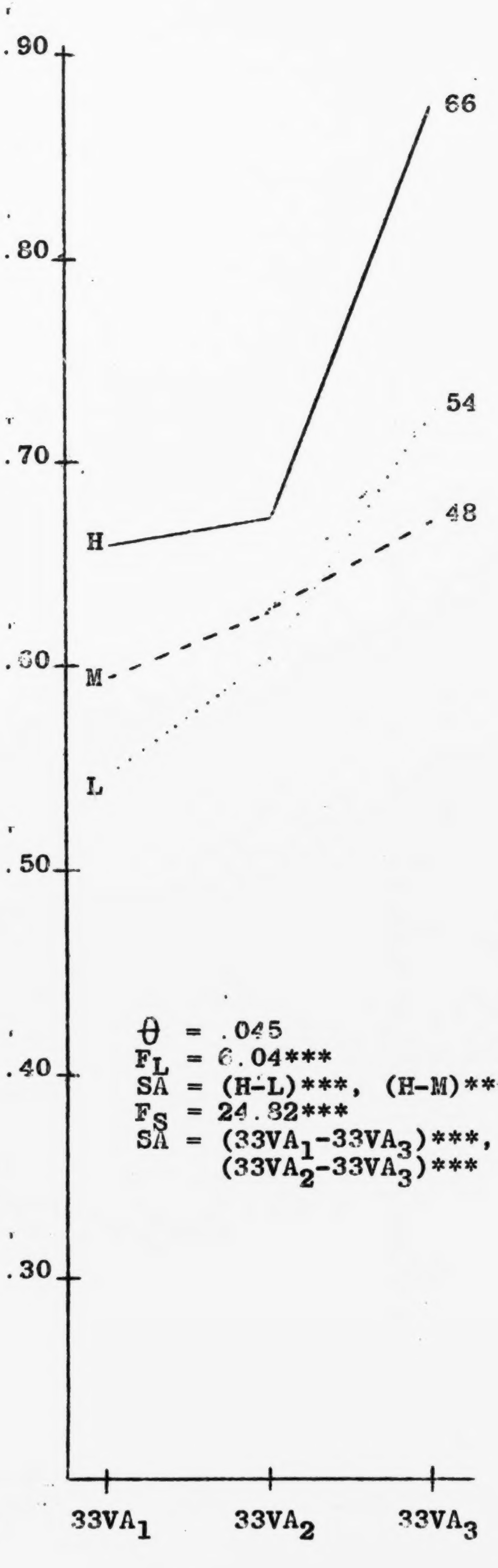

.80
.70
.60
.50
.40
.30
.20
score
H
M
L
103
87
96
$\Theta = .028$
$F_L = 10.78$***
SA = (H-M)**,
(H-L)***
$F_S = 15.70$***
SA = (31-35)***,
(33-35)***
31VA$_2$
33VA$_2$
35VA$_2$
.80
.70
.60
.50
.40
.30
.20
$\Theta = .008$
$F_L = 20.93$***
SA = (H-M)***,
(H-L)***
$F_S = 19.46$***
SA = (31-35)***,
(33-35)***
H
M
L
102
86
93
31VB$_2$
33VB$_2$
35VB$_2$

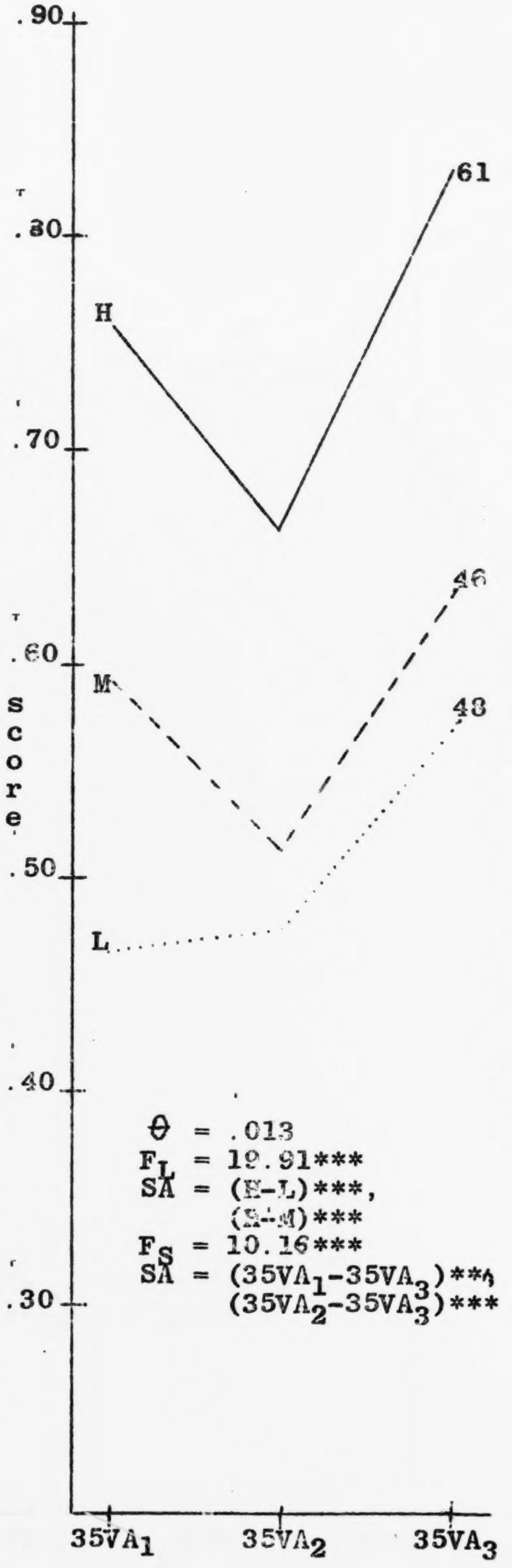

.90
.80
.70
.60
.50
.40
.30
score
H
L
M
9
12
19
θ = .10
F_L = 2.98*
SA = (H-L)*
F_S = 1.25
SA = NS
31P3
33P3
35P3
.90
.80
.70
.60
.50
.40
.30
H
M
L
129
99
112
θ = .025
F_L = 24.35***
SA = (H-M)** (H-L),
(M-L) all ***
F_S = 42.62***
SA = (31-35)***,
(33-35)***
31VA3
33VA3
35VA3

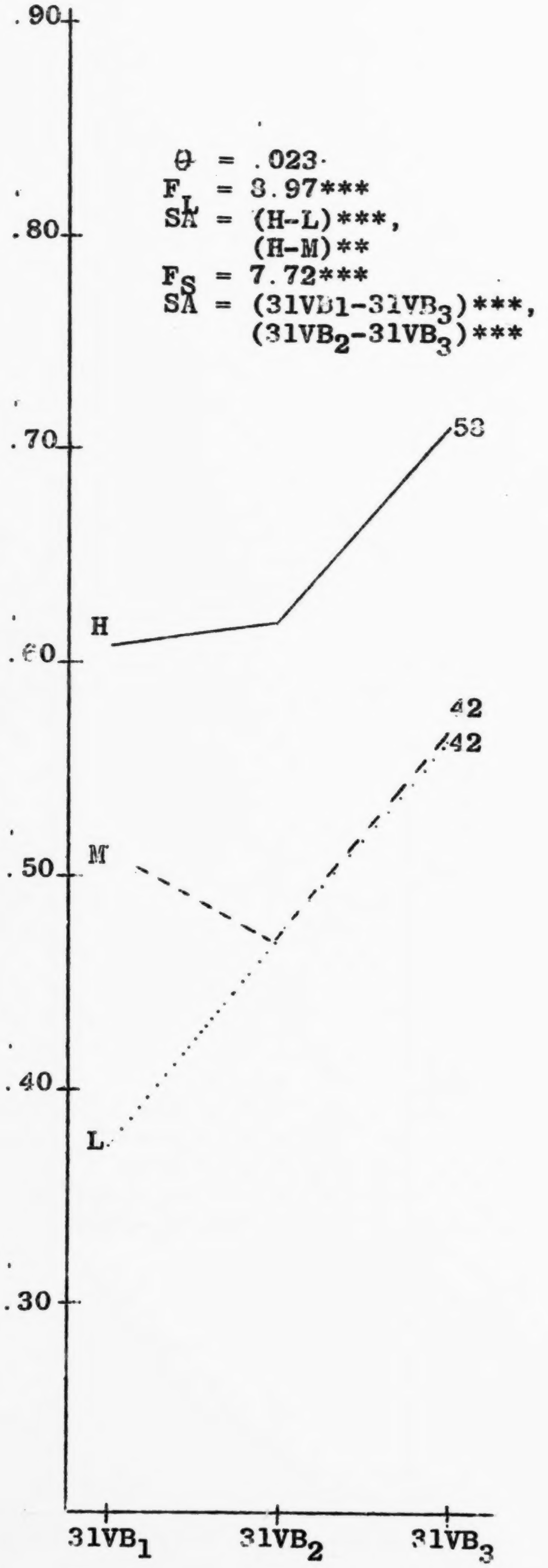

.80
.70
.60
.50
.40
.30
.20
score
θ = .005
F_L = 14.61***
SA = (H-L)***,
(H-M)*, (M-L)**
F_S = 25.98***
SA = (31-35)***
(33-35)***
H
M
L
123
93
101
31VB3
33VB3
35VB3
.80
.70
.60
.50
.40
.30
.20
M
H
L
9
19
11
θ = .140
F_L = .894
SA = NS
F_S = 4.08**
SA = (31P2-31P3)*
31P1
31P2
31P3

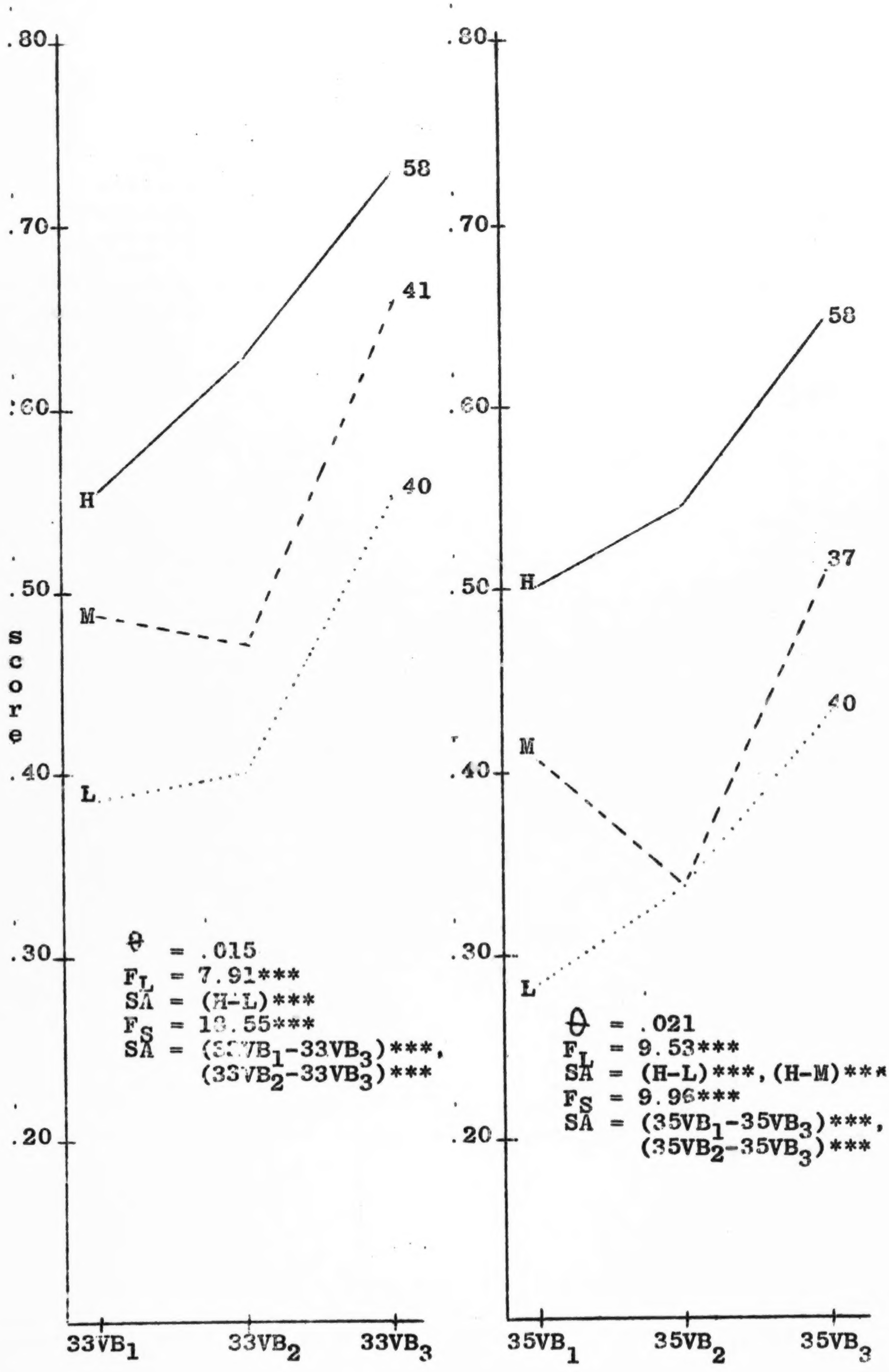

.80
.70
.60
.50
.40
.30
.20
score
M
H
L
9
10
19
θ = .033
F_L = 1.58
SA = NS
F_S = 15.63***
SA = ($33P_1$-$33P_3$)***,
($33P_2$-$33P_3$)***
$33P_1$
$33P_2$
$33P_3$
θ = .048**
F_L = 3.16*
SA = (H-L)*
F_S = 9.15***
SA = ($35P_1$-$35P_3$)***
($35P_2$-$35P_3$)***
8
8
18
H
M
L
$35P_1$
$35P_2$
$35P_3$


.90
.80
.70
.60
.50
.40
.30
score
H
M
L
67
54
59
θ = .027
F_L = 8.38***
SA = (H-L)***,
(M-L)***
F_S = 34.43***
SA = ($31VA_1$-$31VA_3$)***
($31VA_2$-$31VA_3$)***
$31VA_1$
$31VA_2$
$31VA_3$


.90
.80
.70
.60
.50
.40
.30
H
M
L
66
54
48
θ = .045
F_L = 6.04***
SA = (H-L)***, (H-M)***
F_S = 24.82***
SA = ($33VA_1$-$33VA_3$)***,
($33VA_2$-$33VA_3$)***
$33VA_1$
$33VA_2$
$33VA_3$


.90
.80
.70
.60
.50
.40
.30
score
H
M
L
61
46
48
θ = .013
F_L = 12.91***
SA = (H-L)***,
(H-M)***
F_S = 10.16***
SA = (35VA_1-35VA_3)***,
(35VA_2-35VA_3)***
35VA_1
35VA_2
35VA_3


.90
.80
.70
.60
.50
.40
.30
θ = .023
F_L = 8.97***
SA = (H-L)***,
(H-M)**
F_S = 7.72***
SA = (31VB1-31VB_3)***,
(31VB_2-31VB_3)***
H
M
L
58
42
42
31VB_1
31VB_2
31VB_3



.80
.70
.60
.50
.40
.30
.20
score
H
M
L
58
41
40
θ = .015
F_L = 7.91***
SA = (H-L)***
F_S = 13.55***
SA = $(33VB_1-33VB_3)$***, $(33VB_2-33VB_3)$***
$33VB_1$
$33VB_2$
$33VB_3$
.80
.70
.60
.50
.40
.30
.20
H
M
L
58
37
40
θ = .021
F_L = 9.53***
SA = (H-L)***, (H-M)***
F_S = 9.96***
SA = $(35VB_1-35VB_3)$***, $(35VB_2-35VB_3)$***
$35VB_1$
$35VB_2$
$35VB_3$